

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18222 A1(51) International Patent Classification⁷: **C12N 15/85**, 15/90, 1/21 // 15/52, 15/52, (C12N 1/21, C12R 1:19)

(21) International Application Number: PCT/US00/22700

(22) International Filing Date: 18 August 2000 (18.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 09/390,479 7 September 1999 (07.09.1999) US

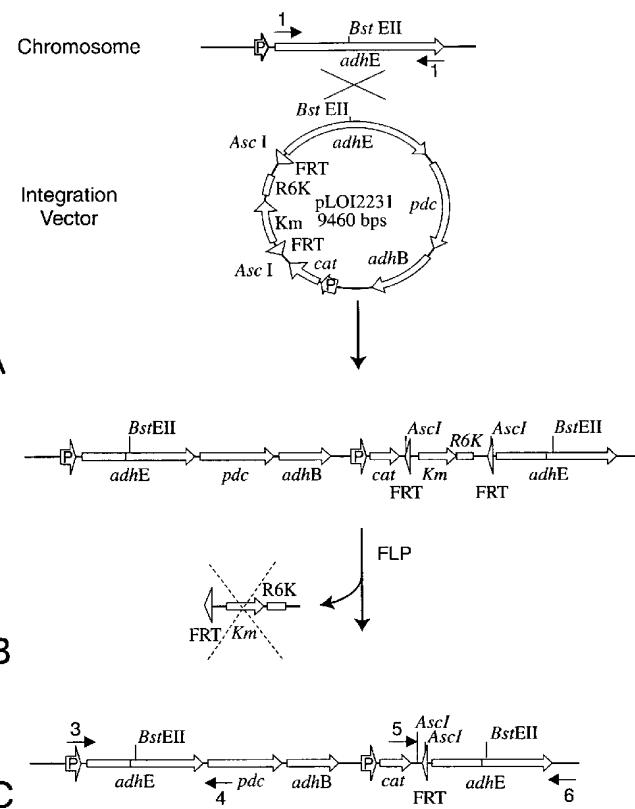
(71) Applicant: **UNIVERSITY OF FLORIDA** [US/US]; 1938 W. University Ave., Gainesville, FL 32611 (US).(72) Inventors: **BORGES, Ana, Claudia**; 3601 SW 31st Drive, Apt. 5D, Gainesville, FL 32608 (US). **ZALDIVAR, Jesus**; 3601 SW 31st Drive, Apt. 5D, Gainesville, FL 32608 (US). **MORALES, Fernando, Martinez**; 720SW 34th Street, Apt. G62, Gainesville, FL 32607 (US). **JIMENEZ, Alfredo, Martinez**; 720 SW 34th St., Apt. K111, Gainesville, FL 32607 (US). **INGRAM, Lonnie, O'Neal**; 3132 NW 57th Terrace, Gainesville, FL 32606 (US).(74) Agents: **DECONTI, Giulio, A., Jr.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR THE CHROMOSOMAL INTEGRATION OF HETEROLOGOUS SEQUENCES

(57) Abstract: The invention provides gene constructs with *in vivo* excisable selection markers, and methods for their use, that are suitable for use in the chromosomal integration of heterologous DNA. Ethanologenic host cells, and methods for engineering such cells, are also provided.**WO 01/18222 A1**



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

Published:

— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS AND COMPOSITIONS FOR THE CHROMOSOMAL INTEGRATION OF HETEROLOGOUS SEQUENCES

Related Information

- 5 The contents of the patents, patent applications, and references cited throughout this specification are hereby incorporated by reference in their entireties.

Government Sponsored Research

- This work was supported, in part, by grants from the Florida Agricultural
10 Experiment Station (publication number R-06853) U.S. Department of Agriculture,
National Research Initiative (98-35504-6177 & 98-35505-6976), and the U.S.
Department of Energy, Office of Basic Energy Science (DE-FG02-96ER20222).

Background of the Invention

- 15 Plasmid vectors are versatile tools which facilitate the isolation, expression and analysis of genes (Bolivar, F., *et al.*, *Gene* 2: 95-113 (1977)). Useful characteristics include the facile production of identical DNA for subsequent *in vitro* and *in vivo* manipulation, the presence of multiple cloning sites (MCS), selectable markers which allow rapid screening for new or improved traits, and the ease with which they can be established as multiple cellular copies to alter gene expression in recombinant hosts.
20 However, the physiological burden imposed by multiples copies of plasmid genes, potential for internal re-arrangements, and segregational instability are disadvantages for many biotechnological applications (Perepelchuk, M. Y., *Gene* 187: 231-238 (1997)).

Antibiotic-resistance genes are frequently used for plasmid maintenance.
25 Alternative selectable markers based on metabolic deficiencies of the host (Degryse, E., *J. Biotech.* 18: 29-40 (1991)) pose further complications for improvement cycles in production strains. For applications such as the deliberate field release, development of organisms for use in food products, and development of biocatalysts for bulk chemicals, special requirements for plasmid maintenance are undesirable.
30 Many of the problems associated with plasmids can be eliminated by the chromosomal integration of desired traits. Integration tools based on modified transposons and conditional plasmid replicons have been developed (de Lorenzo, V., *et*

- 2 -

- al., J. Bacteriol.* 172: 6568-6572 (1990); de Lorenzo, V., *et al.*, *Methods Enzymol.* 235: 386-405 (1994); Rode, C. K., *et al.*, *Gene*. 166: 1-9 (1995); Hamilton, C. H., *et al.*, *J. Bacteriol.* 171: 4617-4622 (1989); Kaniga, K., *et al.*, *J. Bacteriol.* 109: 137-141 (1991); Le Borgne, S., *et al.*, *Gene* 223: 213-219 (1998); Link, A.J., *et al.*, *J. Bacteriol.* 179: 5 6228-6237 (1997)). With these tools, integration can be random or precisely directed by DNA fragments homologous to the host genome. However, complications still remain with most integration systems such as the persistence of selectable markers, transposons, or replicons. For strains in which multiple alterations or continuing improvements are desired, the accumulation of markers and delivery systems can be troublesome.
- 10 Selectable events may be limited by the availability of functional markers. Integrated DNA (replicons, transposon genes, and selectable marker genes) can serve as a site for homologous recombination events which interfere with targeting or randomness during subsequent constructions. Also, the persistence of replicons and transposons increase the potential for gene transfer to other organisms in the environment.
- 15 Replicons and transposons can be eliminated by transforming with purified DNA fragments which lack replication functions (Hasan, N., *et al.*, *Gene* 150: 51-56 (1994); Ohta, K., *et al.*, *Appl. Environ. Microbiol.* 57: 893-900 (1991)). Non-antibiotic markers are available but are often less efficient than antibiotics (de Lorenzo, V., *et al.*, *J. Bacteriol.* 172: 6568-6572 (1990); de Lorenzo, V., *et al.*, *Methods Enzymol.* 235: 386-20 405 (1994); Herrero, M., *et al.*, *J. Bacteriol.* 172: 6557-6567 (1990)). In a few cases, loss of functions such as tetracycline-sensitivity and absence of sucrose-*sacB* system can be selected directly (Bochner, B. R., *et al.* *J. Bacteriol.* 143: 926-933 (1980); Kaniga, K., *et al.*, *J. Bacteriol.* 109: 137-141 (1991); Ried, J.L., *et al.*, *Gene* 57: 239-246 (1985)). However, loss of function due to a mutation is typically not a precise event and 25 can result from unstable point mutations, partial deletion of the resistance gene, or extended deletions which impair the host.

Summary of the Invention

- The foregoing limitations are overcome using the method and vectors of the 30 present invention. In particular, the invention provides a method for integrating nucleic acids into a genome in such a way that any unwanted vector or selectable marker DNA can be removed. This allows for the genome of the recipient host cell to be made

- 3 -

substantially free of any unnecessary nucleic acid, *e.g.*, vector sequence, marker sequence, that can lead to genomic instability.

Accordingly, in one aspect, the invention provides a method for integrating a nucleic acid construct into the genome of a host cell by contacting the cell with a nucleic acid construct under conditions such that the nucleic acid construct is integrated by the cell. The method uses a nucleic acid construct that contains a passenger sequence and a marker sequence, where the marker sequence is flanked by a first and second recombining site. In one embodiment, the first and second recombining site are oriented in the same direction.

10 In another embodiment, the method employs a construct that further contains an origin of replication between the first and second recombining sites, *e.g.*, a conditional origin of replication (*i.e.*, replicon) preferably, *e.g.*, pSC101ori, R6K- γ ori, colE1, oriEV, or an origin of replication derived therefrom.

15 In another embodiment, the nucleic acid construct of the above aspect contains a sequence that contains a promoter, a restriction site, an intron, an exon, an IRES element, a polyadenylation site, or a combination thereof.

In yet another related embodiment, the nucleic acid construct of the above aspect contains a guide sequence capable of directing site-specific integration of the nucleic acid construct to a specific site in the sequence of a replicating genome.

20 In even another embodiment, the above method involves exposing the targeted cell to a site-specific recombinase that results in recombination between the first and second recombining sites of the foregoing nucleic acid construct such that the intervening sequence flanked by the recombining sites is excised. The method involves using a recombinase such as Xer, Int, Cre, and preferably, FLP recombinase.

25 Accordingly, in a related embodiment, the corresponding recombining sites are dif, att, loxP, and preferably FRT. In a preferred embodiment, the above method can be repeated such that the sequential introduction of more than one genetic element may be introduced.

30 In still another embodiment, the passenger sequence encodes at least one gene, preferably a gene involved in ethanogenesis, such as, for example, *adh* or *pdc*. In a related embodiment, the gene involved in ethanogenesis may be derived from a

- 4 -

prokaryote or a eukaryote. In another related embodiment, the passenger sequence may further contain a promoter 5' to the passenger sequence.

In another embodiment, the above method employs a nucleic acid construct containing a marker sequence that encodes a selectable gene, *e.g.*, an antibiotic 5 resistance gene or a non-antibiotic resistance gene. In a preferred embodiment, the antibiotic resistance gene is the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, or chloramphenicol resistance gene. In another preferred embodiment, the non-antibiotic resistance gene is an auxotrophic gene, a metal ion resistance gene, a *trp* gene, or a gene 10 encoding a green fluorescent protein.

In a preferred embodiment of the first aspect, the method employs a bacterial cell, preferably a Gram-negative bacterial cell, more preferably a facultatively anaerobic bacterial cell, more preferably a bacterial cell selected from the family Enterobacteriaceae, and most preferably, a bacterial cell of the genus *Klebsiella* or 15 *Escherichia* (*e.g.*, *E. coli* B or *E. coli* K12). In a related embodiment, the host cell is a recombinant bacterial cell. In another related embodiment, the method uses a nucleic acid construct that is, or is derived from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225, pLOI2226, pLOI2227, pLOI2228, and pLOI2403.

In a second aspect, the invention provides a method for producing a recombinant 20 ethanologenic cell by contacting a cell with a nucleic acid construct under conditions in which integration of the nucleic acid construct occurs resulting in the formation of a recombinant ethanologenic cell. The method uses a nucleic acid construct that contains a passenger sequence that contains an ethanologenic gene, and a marker sequence, flanked by a first and second recombining site. In one embodiment, the first and second 25 recombining site are oriented in the same direction.

In another embodiment, the passenger sequence encodes an ethanologenic gene such as *adh* or *pdc*. In a related embodiment, the passenger sequence encodes *pdc*, *adhB*, and *cat*.

In another embodiment, the nucleic acid further contains a guide sequence 30 thereby resulting in site-specific integration of the nucleic acid construct. In a related embodiment, the guide sequence is derived from a replicating genome.

- 5 -

In yet another embodiment, the method employs a construct that further contains an origin of replication between the first and second recombining sites, *e.g.*, a conditional origin of replication (*i.e.*, replicon) preferably, *e.g.*, pSC101ori, R6K- γ ori, colE1, oriEV, or an origin of replication derived therefrom.

5 In even another embodiment, the above method involves exposing the targeted cell with a site-specific recombinase that results in recombination between the first and second recombining sites of the foregoing nucleic acid construct such that the intervening sequence flanked by the recombining sites is excised. The method involves using a recombinase such as Xer, Int, Cre, and preferably, FLP recombinase.

10 Accordingly, in a related embodiment, the corresponding recombining sites are dif, att, loxP, and preferably FRT.

In still another embodiment, the above method employs a nucleic acid construct containing a marker sequence that encodes a selectable gene, *e.g.*, an antibiotic resistance gene or a non-antibiotic resistance gene. In a preferred embodiment, the 15 antibiotic resistance gene is the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, or chloramphenicol resistance gene. In another preferred embodiment, the non-antibiotic resistance gene is an auxotrophic gene, a metal ion resistance gene, a *trp* gene, or a gene encoding a green fluorescent protein.

20 In a preferred embodiment of the foregoing aspect, a recombinant ethanologenic cell is produced according to the above method.

In a third aspect, the invention provides a recombinant host cell having a nucleic acid construct that contains a passenger sequence and a marker sequence, flanked by a first and second recombining site. In one embodiment, the first and second recombining site are oriented in the same direction.

25 In a related embodiment, the host cell contains a nucleic acid construct where the passenger sequence encodes a gene involved in ethanogenesis such as, *e.g.*, *adh* or *pdc*. In a related embodiment, the passenger sequence is selected from the group including *adh*, *pdc*, and *cat*.

- 6 -

In one embodiment, the host cell contains a nucleic acid construct that further contains an origin of replication between the first and second recombining sites, *e.g.*, a conditional origin of replication (*i.e.*, replicon) preferably, *e.g.*, pSC101ori, R6K- γ ori, colE1, oriEV, or a an origin of replication derived therefrom.

5 In another embodiment, the host cell contains a nucleic acid construct containing a guide sequence capable of directing site-specific integration of the nucleic acid construct to a specific site in the sequence of a replicating genome.

In yet another embodiment, the host cell contains a nucleic acid construct containing a marker sequence that encodes a selectable gene, *e.g.*, an antibiotic 10 resistance gene or a non-antibiotic resistance gene. In a preferred embodiment, the antibiotic resistance gene is the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, or chloramphenicol resistance gene. In another preferred embodiment, the non-antibiotic resistance gene is an auxotrophic gene, a metal ion resistance gene, a *trp* gene, or a gene 15 encoding a green fluorescent protein.

In another embodiment, the host cell is exposed to a site-specific recombinase that results in recombination between the first and second recombining sites of the foregoing nucleic acid construct such that the intervening sequence flanked by the recombining sites, is excised. In a preferred embodiment, the recombination involves 20 using a recombinase such as Xer, Int, Cre, and preferably, FLP recombinase. Accordingly, in a related embodiment, the corresponding recombining sites are dif, att, loxP, and preferably FRT.

In a preferred embodiment, the host cell is a bacterial cell, *e.g.*, a recombinant cell and/or an ethanologenic cell, preferably a Gram-negative bacterial cell, more 25 preferably a facultatively anaerobic bacterial cell, more preferably selected from the family Enterobacteriaceae, and most preferably, of the genus *Klebsiella* or *Escherichia* (*e.g.*, *E. coli* B or *E. coli* K12). In a related embodiment, the host cell contains, a nucleic acid construct that is, or is derived from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225, pLOI2226, pLOI2227, pLOI2228, and pLOI2403.

30 In a fourth aspect, the invention provides a method for producing ethanol by providing a recombinant ethanologenic cell with a nucleic acid construct that contains a passenger sequence and a marker sequence, where the marker sequence is flanked by a

- 7 -

first and second recombining site. The method includes contacting the cell with a substrate which can be fermented into ethanol, such that expression of the passenger sequence results in the production of ethanol.

In one embodiment, the first and second recombining site are oriented in the
5 same direction.

In another embodiment, the passenger sequence encodes a gene involved in ethanogenesis, such as, *e.g.*, *adh* or *pdc*. In a related embodiment, the passenger sequence encodes *adhB*, *pdc*, and *cat*.

In another embodiment, the method employs a construct that further contains an
10 origin of replication between the first and second recombining sites, *e.g.*, a conditional origin of replication (*i.e.*, replicon) preferably, *e.g.*, pSC101ori, R6K- γ ori, colE1, oriEV, or a an origin of replication derived therefrom.

In yet another embodiment, the nucleic acid construct of the above aspect
15 contains a guide sequence capable of directing site-specific integration of the nucleic acid construct to a specific site in the sequence of a replicating genome.

In yet another embodiment, the above method involves exposing the targeted cell with a site-specific recombinase that results in recombination between the first and second recombining sites of the foregoing nucleic acid construct such that the intervening sequence flanked by the recombining sites is excised. The method involves
20 using a recombinase such as Xer, Int, Cre, and preferably, FLP recombinase. Accordingly, in a related embodiment, the corresponding recombining sites are dif, att, loxP, and preferably FRT.

In still another embodiment, the above method employs a nucleic acid construct containing a marker sequence that encodes a selectable gene, *e.g.*, an antibiotic
25 resistance gene or a non-antibiotic resistance gene. In a preferred embodiment, the antibiotic resistance gene is the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, or chloramphenicol resistance gene. In another preferred embodiment, the non-antibiotic resistance gene is an auxotrophic gene, a metal ion resistance gene, a *trp* gene, or a gene
30 encoding a green fluorescent protein.

In a preferred embodiment, the method employs a host cell that a bacterial cell, e.g., a recombinant bacterial cell, preferably a Gram-negative bacterial cell, more preferably a facultatively anaerobic bacterial cell, more preferably selected from the family Enterobacteriaceae, and most preferably, either *Klebsiella* or *Escherichia* (e.g., *E. coli* B or *E. coli* K12).

In one embodiment, the method uses a nucleic acid construct that is, or is derived from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225, pLOI2226, pLOI2227, pLOI2228, and pLOI2403.

In a fifth aspect, the invention provides a nucleic acid construct containing a passenger sequence, and a marker sequence, where the marker sequence is flanked by a first and second recombining site. In a related embodiment the first and second recombining site may be oriented in the same direction.

In another embodiment, the construct further contains an origin of replication between the first and second recombining sites, e.g., a conditional origin of replication (i.e., replicon) preferably, e.g., pSC101ori, R6K- γ ori, colE1, oriEV, or a an origin of replication derived therefrom.

In another embodiment, the nucleic acid construct contains a guide sequence derived from a replicating genome. In a related embodiment, the guide sequence is derived from a bacterial cell.

In yet another embodiment, the construct further contains at least one unique restriction enzyme site. In yet another related embodiment, the passenger sequence of the construct contains an ethanologenic gene, such as, e.g., *adh* or *pdc*, and preferably *adhB*, *pdc*, *cat* or a combination thereof.

In still another related embodiment, the passenger sequence of the construct contains a sequence selected from the group consisting of a heterologous promoter and a prokaryotic termination sequence.

In even another embodiment, the nucleic acid construct contains a marker sequence that encodes a selectable gene, e.g., an antibiotic resistance gene or a non-antibiotic resistance gene. In a preferred embodiment, the antibiotic resistance gene is the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, or chloramphenicol resistance gene. In another preferred embodiment, the non-antibiotic resistance gene is an

- 9 -

auxotrophic gene, a metal ion resistance gene, a *trp* gene, or a gene encoding a green fluorescent protein.

In another embodiment, the construct is exposed to a site-specific recombinase that results in recombination between the first and second recombining sites of the 5 nucleic acid construct such that the intervening sequence flanked by the recombining sites is excised. The excision involves using a recombinase such as Xer, Int, Cre, and preferably, FLP recombinase. Accordingly, in a related embodiment, the construct contains recombining sites such as dif, att, loxP, and preferably FRT.

In a preferred embodiment of the above aspect, the construct is, or is derived 10 from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225, pLOI2226, pLOI2227, pLOI2228, and pLOI2403.

In a related embodiment, the invention provides a kit including at least one of the foregoing nucleic acid constructs and instructions for use.

15 Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 shows a schematic of various integration vectors and helper plasmids 20 (see text for details).

Figure 2 shows a schematic illustrating the use of an integration vector for the insertion of heterologous “passenger” genes into a host genome. A “guide” sequence allows for the site specific targeting of the passenger sequence via homologous recombination. A helper plasmid provides a recombinase (FLP) that catalyzes the *in* 25 *vivo* excision of any unnecessary sequence (*e.g.*, the replicon and marker sequence) flanked by recombining sites (FRT) leaving resident in the genome the passenger gene.

Figure 3 shows a diagram illustrating the targeting of heterologous “passenger” genes to a site specific region of a genome (*i.e.*, the *adhE* gene (shaded)) and recombinase-mediated deletion of the replicon and marker used during plasmid 30 construction and initial integration. The alignment of the guide sequence of the targeting vector and the genomic integration site (Panel A), cross-over event (Panel B),

- 10 -

and resultant recombinase-mediated (FLP) excision of the replicon and selectable marker (Panel C) are shown.

Figure 4 is a photograph of an agarose gel stained with ethidium bromide showing PCR amplified nucleic acid fragments amplified using different primers (Panel 5 A) and restriction enzyme digested (Panel B) to confirm the correct chromosomal integration of a heterologous nucleic acid (see text for details).

Detailed Description of the Invention

In order for the full scope of the invention to be clearly understood, the following 10 definitions are provided.

I. Definitions

As used herein the term “host cell” and “recombinant host cell” is intended to include a cell suitable for genetic manipulation, *e.g.*, which can incorporate heterologous 15 polynucleotide sequences, *e.g.*, which can be transfected. The cell can be a microorganism or a higher eukaryotic cell, such as an animal cell or a plant cell. The term is intended to include progeny of the cell originally transfected. In preferred embodiments, the cell is a bacterial cell, *e.g.*, a Gram-negative bacterial cell, and this term is intended to include all facultatively anaerobic Gram-negative cells of the family 20 Enterobacteriaceae such as *Escherichia*, *Shigella*, *Citrobacter*, *Salmonella*, *Klebsiella*, *Enterobacter*, *Erwinia*, *Kluvera*, *Serratia*, *Cedecea*, *Morganella*, *Hafnia*, *Edwardsiella*, *Providencia*, *Proteus*, and *Yersinia*. Particularly preferred recombinant hosts are *Escherichia coli* or *Klebsiella oxytoca* cells. Preferably, the term recombinant host cell is intended to include a cell that has already been selected or engineered to have certain 25 desirable properties and suitable for further modification using the compositions and methods of the invention.

The term “passenger sequence” is intended to include any desired sequence that is intended for integration into the host cell. For example, the passenger sequence may include, a restriction site, multiple restriction sites (*e.g.*, a polylinker or multiple cloning 30 site (MCS)), a unique stretch of sequence suitable for marking the cell as distinct from a wild type cell, a regulatory element, *e.g.*, a polyadenylation sequence, a promoter, an intron, a splice signal, an internal ribosomal entry site (IRES); for regulating the

- 11 -

expression of multiple genes or cistrons, and/or a gene, for example a gene encoding a polypeptide. A passenger sequence encoding a gene may further include a promoter if appropriate or use the endogenous promoter found proximal to the site of integration. In addition, the passenger sequence encoding a gene may further include other regulatory elements, if appropriate, to improve expression. A passenger sequence may comprise genetic elements derived from any source, *e.g.*, eukaryotes, prokaryotes, virii, or synthetic polynucleotide fragments.

The term "guide sequence" is intended to include a sequence that can be located 5', 3', or internal to the sequence intended to be integrated, such that recombination 10 between the introduced vector (or portion thereof) and the recipient host cell genome is accomplished. Typically, a guide sequence has high similarity or identity with a site specific region of the recipient host cell genome such that a targeted integration of the passenger sequence to this site by homologous recombination is accomplished.

The term "site-specific integration" is meant to refer to the integration of an 15 exogenous nucleic acid sequence to a specific area of the genome of a recipient host cell. In general, a guide sequence allows for the homologous recombination between a portion of the targeting vector and the host cell such that a predictable targeting of the sequence to that area of the genome represented in part by the guide sequence, is specifically targeted.

20 The term "marker sequence" is intended to include any sequence that can be encoded by a nucleic acid, introduced into the replicating genome of a recipient host cell, and detected, thereby indicating that the cell has been "marked" by such a sequence. Accordingly, the term is intended to include a sequence for example having a restriction enzyme sequence that can be detected by a corresponding restriction enzyme.

25 In addition, or alternatively, the sequence may be detected using any of a variety of art recognized techniques, *e.g.*, polymerase chain reaction using appropriate primers, nucleic acid hybridization, *etc.* Preferably, the marker sequence encodes a gene product that confers on the cell a selectable phenotype, *e.g.*, resistance to an antibiotic or other cytotoxic agent or a conditional growth advantage. Accordingly, a marker sequence can 30 encode, *e.g.*, the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, chloramphenicol resistance gene, hygromycin resistance gene, thymidine kinase, an

- 12 -

auxotrophic gene, a metal ion resistance gene, a *trp* gene, a gene producing a visual phenotype (*e.g.*, green fluorescent protein), or a gene producing a cell surface antigen (*e.g.*, CD4). Any of the foregoing markers when expressed in cells can be detected using art recognized techniques such as, *e.g.*, appropriate culture conditions or selection 5 techniques (*e.g.*, antibodies, FACS (fluoroscein activated cell sorting), or flow cytometric analysis).

The term “origin of replication” or “replicon” is intended to include any sequence, conditional or otherwise, that can confer on a nucleic acid sequence the ability to be replicated in a cell for the purposes of, *e.g.*, propagating the nucleic acid or 10 maintaining the presence of the nucleic acid in a cell. Such sequences are known in the art and are routinely incorporated into the backbone of many nucleic acid vectors to facilitate their propagation and typically include, *e.g.*, pSC101ori, R6K-*yori*, colE1ori, or oriEV.

The term “recombinase site” is intended to include a nucleic acid element that 15 represents a site specific binding site for a recombinase. Examples of such elements include, *e.g.*, the FRT sequence, the dif sequence, the att sequence, and the loxP sequence.

The term “recombinase” is intended to include any recombinase having a site specific recombinase activity. Examples of such recombinases are FLP, Xer, Int, and 20 Cre and these enzymes typically have site specific recombinase activity on, respectively, the FRT sequence, the dif sequence, the att sequence and the loxP sequence.

The term “gene involved in ethanogenesis” is intended to include any gene capable of conferring on a cell ethanogenic properties or capable of improving any aspect of cellular ethanogenesis, such as, *e.g.*, substrate uptake, substrate processing, 25 ethanol tolerance, *etc.* Genes involved in ethanogenesis are, *e.g.*, alcohol dehydrogenase, pyruvate decarboxylase, secretory protein/s, and polysaccharases, and these genes, or their homologs, may be derived from any appropriate organism.

The term “substrate” is intended to include any moiety that can be converted into ethanol. Substrates that are suitable for converting into ethanol include sugar moieties 30 such as, *e.g.*, monosaccharides, disaccharides, trisaccharides, oligosaccharides, and complex carbohydrates, including, *e.g.*, lignocellulose, which comprises cellulose, hemicellulose, and pectin.

- 13 -

The term "derived from" is intended to include the isolation (in whole or in part) of a polynucleotide segment from an indicated source. The term is intended to include, for example, direct cloning, PCR amplification, or artificial synthesis from, or based on, a sequence associated with the indicated polynucleotide source.

- 5 The term "ethanologenic" is intended to include the ability of a microorganism to produce ethanol from a carbohydrate as a primary fermentation product. The term is intended to include naturally occurring ethanologenic organisms, ethanologenic organisms with naturally occurring or induced mutations, and ethanologenic organisms which have been genetically modified.
- 10 The term "Gram-negative bacteria" is intended to include the art recognized definition of this term. Typically, Gram-negative bacteria include, for example, the family Enterobacteriaceae which comprises, among others, the species *Escherichia* and *Klebsiella*.

15 **II. Nucleic Acid Constructs**

The invention provides a number of novel nucleic acid constructs that are suitable for targeting a heterologous nucleic acid sequence to the genome or chromosome of a recipient host cell.

- 20 The heterologous sequence or passenger sequence of the targeting vector can be any desirable sequence that is useful when introduced into a cell. Accordingly, the sequence may be a regulatory element (e.g., a promoter, intron, splicing signal, internal ribosome entry site, polyadenylation signal) or a gene encoding a gene product, e.g., a polypeptide, or a combination thereof (e.g., a gene encoding a polypeptide that is under the transcriptional control of a promoter). In addition, the heterologous sequence may
- 25 be targeted to any replicating genome, e.g. a bacterial cell, a yeast, an insect cell, a plant cell, or an animal cell. It is well known in the art as to what regulatory elements, e.g., promoters, etc., are suitable for use in, e.g., a bacterial cell versus a plant or animal cell.

- If site specific integration of the passenger sequence is desired (as opposed to random integration) a guide sequence must be incorporated into the nucleic acid vector.
- 30 Typically the guide sequence contains a portion of the genomic locus to be targeted. Considered design of this region allows for the accurate placement of the passenger

- 14 -

sequence anywhere in the genome, preferably, *e.g.*, under the control of a desirable endogenous promoter.

To facilitate the construction of the targeting vector, an origin of replication, or replicon, and a marker sequence is included in the vector. Specifically exemplified in 5 the present invention are the markers encoding antibiotic resistance genes to ampicillin, kanamycin, and chloramphenicol, however it will be appreciated that other selectable markers may be used as described above. In part, the selection of an appropriate marker will depend on the organism to be targeted, *e.g.*, prokaryotic versus eukaryotic. The characteristics of the foregoing markers and their efficacies in various organisms is well 10 known in the art.

In addition to an appropriate marker, an origin of replication or replicon is also incorporated into the targeting construct to facilitate propagation of the vector during the construction and/or initial integration of the vector. Specifically exemplified are the pSC101 and R6K- γ replicons however other replicons, *e.g.*, colE1, and oriEV as well as 15 replicons suitable for use in eukaryotic cells are also encompassed by the invention.

Importantly, the constructs of the invention possess site specific recombining sites that flank all of the vector sequence that, after targeting of the passenger sequence to the genome, is no longer necessary or desired. Indeed, one advantage of the invention is that following targeting and integration of the passenger sequence to the genome of 20 the host cell, the *in vivo* excision of any undesired sequence can be accomplished using a recombinase that binds the site specific recombining sites contained in the integrated vector.

Critical for allowing the correct excision of the undesired vector sequence is the placement and directionality of the recombining sites. The novel vectors of the 25 invention possesses a first and a second recombining site that flank all of the replicon and marker sequence and are oriented in the same direction. This allows for the complete excision of all the vector sequence leaving behind the desired targeted passenger sequence and a minimal remainder of the joined recombining sites (68 bps).

While the FRT recombining site and corresponding FLP recombinase is 30 exemplified (Senecoff, J., *et al.* P.N.A.S. 82:7270-7274 (1985); Szybalski, W., *Gene* 135:279-290 (1993); Hoang, T.T., *et al.*, *Gene* 212:77-86 (1998)), a number of other site specific recombining sites (*e.g.*, dif, att, loxP) and corresponding recombinases (*e.g.*,

- 15 -

Xer, Int, Cre) may be used. Accordingly, any of the above described markers may be used after the excision step because the recipient host cell has once again been rendered sensitive to the selectable marker. Thus, one can sequentially introduce multiple genes into an organism without ever running into the difficulty of having to change markers or
5 exhaust all suitable markers. Accordingly, if a particular selectable marker works efficiently in a particular organism, it may be exclusively used to introduce multiple passenger sequences into a target organism.

In a preferred embodiment, the constructs of the invention are modified to contain a passenger sequence encoding ethanol pathway gene products such as, e.g.,
10 alcohol dehydrogenase (e.g., *adhB*) and pyruvate decarboxylase (*pdc*), that are necessary for ethanol production in, e.g., a microorganism. A guide sequence specific for the endogenous gene in the target organism allows for the targeting of these genes to a specific locus such that the heterologous ethanol pathway genes (i.e., *adhB* and *pdc*) are brought under the transcriptional control of an appropriate endogenous promoter (e.g.,
15 *adhE*).

Accordingly, the foregoing novel constructs allow for the genetic engineering of superior host cells suitable, e.g., for various industrial applications, such as, e.g., use in a fermentation reaction for producing, e.g., ethanol (for a review of other industrial applications, see, e.g., Barrios-Gonzalez *et al.*, *Biotechnol. Ann. Rev.* 2:85-121 (1996);
20 *From Ethnomycology to Fungal Biotechnology: Exploiting from Natural Resources for Novel Products*, Singh, J., Ed., Plenum Press, Pub. (1999); *Manual of Industrial Microbiology and Biotechnology*, Demain, A. Ed., Am. Soc. of Microbiology, Pub., (1999); *Biomining: Theory, Microbes, and Industrial Processes*, Rawlings, Ed., R.G. Landes Co., Pub. (1997); *Biotechnology of Industrial Antibiotics*, Vandamme, E., Ed.,
25 Marcel Dekker, Pub. (1984); *Industrial Biotechnology*, Malik, V., Ed., Science, Pub. (1992); *Biotechnology and Food Ingredients*, Goldberg *et al.*, Ed., Aspen Publishers (1991); *Biotechnology and Food Process Engineering*, Schwartzberg *et al.*, Ed., Marcel Dekker, Pub.(1990); and *Food Biotechnology: Techniques and Applications*, Mittal, G., Technomic Pub. Co. (1992).

III. Methods of Use

In a preferred embodiment, the methods of the invention apply the constructs described above for the integration of one or more desirable gene sequences into a host cell. Accordingly, the methods of the invention have immediate application in the 5 genetic engineering of, e.g., microorganisms (e.g., for various industrial applications), plants (e.g., for pest resistance, hardiness, yields), and animal cells (e.g., for producing cytokines, hormones, etc.).

In one particular application of the methods of the invention, one or more genes necessary for producing ethanol are provided on a plasmid or integrated into a host 10 chromosome using the method of the invention. More preferably, essential genes for fermenting a sugar substrate into ethanol, e.g., pyruvate decarboxylase (e.g., *pdc*) and/or alcohol dehydrogenase (e.g., *adh*, preferably *adhB*) are introduced into the host of the invention using an bicistronic operon or PET operon as described in U.S.P.N. 5,821,093, hereby incorporated by reference.

15 Art recognized techniques for introducing nucleic acids into prokaryotic or eukaryotic cells such as e.g., bacteria, yeast, insects, plants, and animals as well as appropriate culturing methods have been described (see, e.g., *Large-Scale Mammalian Cell Culture Technology*, Lubiniecki, A., Ed., Marcel Dekker, Pub., (1990), *Bacterial Cell Culture: Essential Data*, Ball, A., John Wiley & Sons, (1997), *Molecular and Cell Biology of Yeasts*, Yarranton et al., Ed., Van Nostrand Reinhold, Pub., (1989); *Yeast Physiology and Biotechnology*, Walker, G., John Wiley & Sons, Pub., (1998); 20 *Baculovirus Expression Protocols*, Richardson, C., Ed., Humana Press, Pub., (1998); *Methods in Plant Molecular Biology: A Laboratory Course Manual*, Maliga, P., Ed., C.S.H.L. Press, Pub., (1995); *Current Protocols in Molecular Biology*, eds. Ausubel et 25 al., John Wiley & Sons (1992), Sambrook, J. et al., *Molecular Cloning: A Laboratory Manual. 2nd, ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989), and *Bergey's Manual of Determinative Bacteriology*, Kreig et al., Williams and Wilkins (1984)).

Accordingly, using the methods of the invention, a single genetic construct can 30 be designed to encode all of the necessary gene products (e.g., a glucanase, an endoglucanase, an exoglucanase, a secretory protein/s, pyruvate decarboxylase, alcohol

- 17 -

dehydrogenase) for imparting to a non-ethanologenic (or weakly ethanologenic organism) all the necessary genes for producing ethanol from a sugar substrate.

Alternatively, the invention allows for more than one gene to be introduced serially without having to change markers. In a preferred embodiment, this allows for
5 the engineering of a host cell suitable for performing simultaneous saccharification and fermentation (SSF). In addition, it will also be appreciated that such a host may be further manipulated, using methods known in the art, to have mutations in any endogenous gene/s (*e.g.*, recombinase genes) that would interfere with the stability, expression, and function of the introduced genes. Further, it will also be appreciated that
10 10 the invention is intended to encompass any regulatory elements, gene/s, or gene products, *i.e.*, polypeptides, that are sufficiently homologous to the ones described herein.

Methods for screening strains having the introduced genes may be performed using, *e.g.*, PCR amplification using site or gene specific primers, visual screens that can
15 15 identify cells expressing either the desired integrated gene, marker, or absence of the marker and any other unnecessary nucleic acid or other art recognized techniques.

For example, for screening an ethanologenic microorganism, the ADH gene product produces acetaldehyde that reacts with the leucosulfonic acid derivative of p-roseaniline to produce an intensely red product. Thus, ADH-positive clones can be
20 20 easily screened and identified as bleeding red colonies.

Rapid evaluations of ethanol producing potential can also be made by testing the speed of red spot development on aldehyde indicator plates (Conway *et al.*, (1987) *J. Bacteriol.* 169:2591-2597). Typically, strains which prove to be efficient in sugar conversion to ethanol can be recognized by the production of red spots on aldehyde
25 25 indicator plates within minutes of transfer.

In a most preferred embodiment of the invention, the methods of the invention allow for the production of a single host cell that is ethanologenic, that is, has all the necessary genes, either naturally occurring or artificially introduced or enhanced (*e.g.*, using a surrogate promoter and/or genes from a different species or strain), such that the
30 30 host cell has the ability to produce and secrete a polysaccharase/s, degrade a complex sugar, and ferment the degraded sugar into ethanol. Accordingly, such a host is suitable for simultaneous saccharification and fermentation. Moreover, this host cell is free of

marker sequences, unnecessary vector sequence, a heterologous replicon, and these sequences can lead to genomic instability in the organism or make the organism less suitable for release or use outside, *e.g.*, a controlled laboratory environment.

In addition, it should be readily apparent to one skilled in the art that the ability
5 conferred by the present invention, to transform genes coding for a protein or an entire metabolic pathway into a host cell without leaving resident in the host genome unnecessary or undesired nucleic acid has wide application. Envisioned in this regard, for example, is the application of the present invention to a variety of situations where regulatory elements and/or genes may be added at will and moreover, deleterious genes
10 deleted, without unnecessarily littering the genome of the target organism with unwanted nucleic acid.

IV. Host Cells

The invention also relates to host cells that are modified using the foregoing
15 methods and/or nucleic acid constructs of the invention. Preferably, any host cell, *e.g.*, animal, plant, or microbe, can be modified according to the methods of the invention. More preferably, the methods and constructs of the invention allow for the engineering of heterologous DNA into the genome of any microorganism suitable, *e.g.*, for an industrial application. Even more preferably, the invention provides for the engineering
20 of an organism useful for the production of ethanol, *e.g.*, by fermentation.

Accordingly, in one embodiment, a host cell of the invention comprises a heterologous polynucleotide segment encoding a polypeptide under the transcriptional control of a naturally occurring promoter in the genome or, if appropriate, under the control of a heterologous promoter.

25 In another embodiment, the host cell prepared according to the method of the invention is exposed to a recombinase such that the *in vivo* excision of any unnecessary nucleic acid flanked by recombining sites is accomplished while the targeted passenger sequence encoding a desirable gene is left behind. Typically, the nucleic acid sequence between the recombining sites contains, for example, a marker sequence and may further
30 contain an origin of replication or replicon. The host cell of the invention is render free of this sequence by being exposed to a site specific recombinase, preferably encoded by an expressible construct, which is introduced into the cell using standard techniques.

- 19 -

- In a preferred embodiment, the resultant recombinant host cell is made substantially free of any nucleic acid flanked by FRT recombining sites using an FLP recombinase. It is understood that the host cell of the invention may also be modified using nucleic acid constructs having recombining sites such as a dif sequence, att 5 sequence, or loxP sequence and made substantially free of any undesired nucleic acid using a corresponding recombinase, such as, e.g., Xer, Int, or Cre.

The resultant host cell as prepared according to the method of the invention has the advantage of no longer having any unnecessary heterologous DNA which can lead to, e.g., genomic stability. In addition, the host cell no longer encodes a marker 10 sequence and thus can be retargeted with the same selectable marker sequence in order to introduce another genetic modification to the cell. Thus, the serial introduction of multiple genes can be achieved using the same selection marker. In theory at least, there appears to be no limit as to the number of gene constructs that can be introduced into a host cell of the invention.

15 In a most preferred embodiment, the host cell of the invention contains at least one passenger sequence containing, e.g., a gene of interest, e.g., a gene encoding a desired polypeptide for use in the bioconversion of a sugar to ethanol, or a step thereof. Preferred ethanologenic genes include those that encode, e.g., any ethanol pathway gene that can facilitate the production of ethanol from the cell or extract thereof, such as 20 alcohol dehydrogenase, pyruvate decarboxylase, a cellulase, or a secretory polypeptide. A desired ethanologenic gene may be derived from another species, e.g., a yeast, an insect, an animal, or a plant. The techniques for introducing and expressing such genes in a recombinant host, are presented in the examples.

For example, a non-ethanologenic host cell may be converted into an 25 ethanologenic host by introducing, for example, ethanologenic genes from an efficient ethanol producer like *Zymomonas mobilis*. This type of genetic engineering, using the constructs and methods of the invention, results in a recombinant host capable of efficiently fermenting sugar into ethanol.

Accordingly, the invention makes use of a non-ethanologenic recombinant host, 30 e.g., *E. coli* strain B or *E. coli* strain K-12. These strains may be used to express a desired polypeptide, e.g., an ethanologenic gene introduced using techniques described herein. In a preferred embodiment of the present invention, the host cell having been

- 20 -

modified using the methods and constructs of the invention, can be applied in degrading or depolymerizing a complex saccharide into monosaccharides. Subsequently, the host cell can catabolize the simpler sugar into ethanol by fermentation. This process of concurrent complex saccharide depolymerization into smaller sugar residues followed 5 by fermentation is referred to as simultaneous saccharification and fermentation (SSF).

In another embodiment, the invention makes use of a recombinant host that is ethanologenic and is improved using the methods of the invention. It is understood that an improvement of an ethanologenic host may include, *e.g.*, increasing the ability of organism to produce ethanol, depolymerize a particular substrate, tolerate a higher 10 ethanol level, *etc.* In a preferred embodiment, the recombinant host is a Gram-negative bacterium. In another embodiment, the recombinant host is from the family Enterobacteriaceae. The ethanologenic hosts of U.S.P.N. 5,821,093, hereby incorporated by reference, for example, are suitable hosts and include, in particular, *E. coli* strains KO4 (ATCC 55123), KO11 (ATCC 55124), and KO12 (ATCC 55125). In 15 addition, the LY01 strain may be employed as described in U.S.S.N. 08/834,900 and this application is hereby incorporated by reference.

It will be appreciated that host cells prepared according to the method of the invention may be used in conjunction with another recombinant host that expresses, yet another desirable polypeptide, *e.g.*, a different passenger gene sequence. In a particular 20 example, a non-ethanologenic host cell may be used in conjunction with an ethanologenic host cell and either one or both of these host cells may be engineered according to the methods of the invention to accomplish, *e.g.*, one step of a multistep process, *e.g.*, the converting of a substrate into ethanol. For example, a non-ethanologenic host may be engineered for carrying out, *e.g.*, the depolymerization of a 25 complex sugar followed by the use of an engineered ethanologenic host for fermenting the depolymerized sugar. Accordingly, it will be appreciated that the host cells of the invention may be used serially or contemporaneously for carrying out a particular process.

This invention is further illustrated by the following examples which should not 30 be construed as limiting.

- 21 -

EXEMPLIFICATION

EXAMPLE I

5 **Construction of Targeting Vectors for Introducing Heterologous DNA that is
Excisable *In Vivo***

In this example, selectable targeting vectors for introducing heterologous DNA into a recipient host cell are described.

10 Throughout the example, the following materials and methods were used unless otherwise stated.

Materials and Methods

The bacterial strains and plasmids used in this example are listed in Table 1,
15 below.

- 22 -

Table 1. Bacterial strains and plasmids

<i>E. coli</i> STRAINS	RELEVANT GENOTYPE	SOURCE or REFERENCE
DH5 α	<i>lacZ</i> Δ M15 <i>recA1 endA1 hsdR17</i> [r _K ⁻ m _K ⁺] <i>supE44</i>	Promega
S17-1	<i>thi pro recA hsdR RP4-2-tet::Mu apha::Tn7 Sp^r Tp^r</i>	deLorenzo <i>et al.</i> , 1990
TOP10F'	F' { <i>lacI</i> ^q Tn10(Tc ^r)} <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1 deoR araD139</i> Δ (<i>ara-leu</i>)7697 <i>galU</i> <i>galK rpsL</i> (St ^r) <i>endA1 nupG</i>	Invitrogen
SE2272	prototrophic strain B, Δ <i>frd102 zjd</i>	K. T. Shanmugan collection
SE2275	prototrophic strain K-12, Δ <i>frd102 zjd</i>	K. T. Shanmugan collection
FM7	SE2272 derivative containing pLOI2231 integrated (Km ^R , Cm ^R)	This study
FM19	SE2275 derivative containing pLOI2231 integrated (Km ^R , Cm ^R)	This study
FM18	FM7 derivative (Km ^S , Cm ^R)	This study
FM20	FM19 derivative (Km ^S , Cm ^R)	This study
LY01	<i>frd pfl::pdc adhB cat alc^R</i>	Yomano <i>et al.</i> , <i>J. Ind.</i> <i>Microbiol.</i> <i>Biotech.</i> 20:132-138 (1998)
PLASMIDS	DESCRIPTION	SOURCE OR REFERENCE

- 23 -

Table 1. Continued

pUC18	colE1ori; Ap ^R	Messing, J., <i>Methods Enzymol.</i> 101:20-78 (1983)
pCR2.1-TOPO	colE1ori; Ap ^R ; Km ^R	Invitrogen
LITMUS38	colE1ori; M13ori; Ap ^R	New England Biolabs, Inc.
pSG76A-K-C	γ-R6Kori; FRT; Ap ^R -Km ^R -Cm ^R	Posfai <i>et al.</i> , 1997
pST76A-K-C	pSC101ori; FRT; Ap ^R -Km ^R -Cm ^R	Posfai <i>et al.</i> , 1997
PFT-A	pSC101ori; FLP gene under control of the tetracycline repressor; Ap ^R	Posfai <i>et al.</i> , 1997
pLOI2216	XmnI/HindIII fragment substitution from pSG76-A into pUC18; FRT1; colE1ori; Ap ^R	This study
pLOI2217	HindIII site deletion of pLOI2216; FRT1; colE1ori; Ap ^R	This study
pLOI2218	EcoRI/PstI fragment deletion of pLOI2217; FRT1; colE1ori; Ap ^R	This study
pLOI2219	EcoRI site substitution of Clal site into PLOI2218; FRT1; colE1ori; Ap ^R	This study
pLOI2220	Scal/EcoRI fragment substitution from pLOI2401 into pLOI2219; FRT1; colE1ori; Ap ^R	This study
pLOI2221	Ascl site substitution of EcoRI site into pLOI2220; FRT1; colE1ori; Ap ^R	This study
pLOI2222	HindIII site deletion of pLOI2221; FRT1; colE1ori; Ap ^R	This study
pLOI2223	EcoRI/SphI fragment substitution from pLOI2222 into pSG76-A; FRT1; FRT2; γ-R6Kori; Ap ^R	This study
pLOI2224	EcoRI/SphI fragment substitution from pLOI2222 into pSG76-K; FRT1; FRT2; γ-R6Kori; Km ^R	This study
pLOI2225	EcoRI/SphI fragment substitution from pLOI2222 into pSG76-C; FRT1; FRT2; γ-R6Kori; Cm ^R	This study

Table 1. Continued

pLOI2226	<i>EcoRI/SphI</i> fragment substitution from pLOI2222 into pST76-A; FRT1; FRT2; pSC101ori; Ap ^R	This study
pLOI2227	<i>EcoRI/SphI</i> fragment substitution from pLOI2222 into pST76-K; FRT1; FRT2; pSC101ori; Km ^R	This study
pLOI2228	<i>EcoRI/SphI</i> fragment substitution from pLOI2222 into pST76-C; FRT1; FRT2; pSC101ori; Cm ^R	This study
pLOI2229	<i>EcoRI</i> fragment from pLOI2408 into pLOI2403; colE1ori; Ap ^R ; Km ^R	This study
pLOI2230	<i>BamHI</i> fragment from pLOI510 (Ohta <i>et al.</i> , 1991) into pLOI2229; colE1ori; Ap ^R ; Km ^R ; Cm ^R	This study
pLOI2231	<i>AscI</i> fragment from pLOI2230 into pLOI2224; γ-R6Kori; Km ^R ; Cm ^R	This study
pLOI2400	<i>AscI</i> site substitution of <i>SphI</i> site into pUC18; colE1ori; Ap ^R	This study
pLOI2401	<i>BglII</i> site substitution of <i>KpnI</i> site into pLOI2400; colE1ori; Ap ^R	This study
pLOI2402	<i>AscI</i> site substitution of <i>ApaI</i> site into LITMUS38; colE1ori; M13ori; Ap ^R	This study
pLOI2403	<i>AscI</i> site substitution of <i>StuI</i> site into pLOI2402; colE1ori; M13ori; Ap ^R	This study
pLOI2408	<i>adhE</i> gene into pCR2.1-TOPo; colE1ori; Ap ^R ; Km ^R	This study
pLOI2413	<i>adhE</i> gene into pLOI2403; colE1ori; M13ori; Ap ^R	This study
pLOI510	promoterless <i>pdc adhB</i> genes from <i>Z. mobilis</i> and <i>cat</i> gene cloned into <i>BamHI</i> site of pUC18; colE1ori; Ap ^R ; Cm ^R	Ohta <i>et al.</i> , 1991

For all plasmid constructions, standard methods were employed (Sambrook, J., *et al. Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)). Reagents used in cloning were molecular biology grade and used as directed by the manufacturers. Restriction enzymes, T4 DNA polymerase, and Klenow polymerase were purchased from New England Biolabs (Beverly, MA). *Taq* PCR MasterKit was purchased from Qiagen (Santa Clarita, CA).

10 Polymerase chain reactions (PCR) were carried out using an Eppendorf Mastercycler (Westbury, NY). Primers were obtained from Genosys Biotechnologies (The Woodlands, TX). PCR products were ligated into pCR2.1-TOPo (Invitrogen, Carlsbad, CA) using topoisomerase. DNA fragments were isolated from gels using a QIAquick

- 25 -

Gel Extraction Kit. DNA fragments were assembled using a Rapid DNA ligation Kit (Boehringer Mannheim Corporation, Indianapolis, IN). Wizard Plus kits (Promega, Madison, WI) were used for plasmid purification. Dideoxy sequencing of plasmids was performed by using fluorescent primers and a LI-COR model 4000L DNA Sequencer 5 (Lincoln, NB) as previously described (Lai, X., et al., *Appl. Environ. Microbiol.* 63:355-363 (1996)). Except where mentioned, all ligation reactions were done with the Rapid DNA Ligation Kit and used to transform *E. coli* DH5 α , followed by selection in LB plates containing ampicillin.

10 *Plasmid Constructions*

Two sets of insertion vectors were constructed using, as an initial vector, the suicide plasmids pSG76 and pST76 series, described by Posfai et al. (1997). These low copy number plasmids contain the conditional replicons R6K- γ and pSC101, respectively. Modifications were executed in several steps so that another recombining 15 site, i.e., FRT site, was introduced into each of original vectors, as well as a more suitable multiple cloning site (MCS). Unless mentioned in the text, all new plasmids are shown in brackets, just after its corresponding description. Complete sequences for the following plasmids have been deposited in GenBank with the following accession numbers: pLOI2223, AF172933; pLOI224, AF172934; pLOI225, AF172935; pLOI226, 20 AF172936; pLOI227, AF172937; pLOI228, AF172938 (see also, respectively, SEQ ID NOS: 1-6).

To facilitate the manipulation of each construct, most of the plasmid manipulations were done in pUC18 (Messing, J., *Methods Enzymol.* 101:20-78 (1983)) and derivatives made herein. An 800 bp *XmnI-HindIII* fragment from pUC18 was 25 exchanged for a fragment comprising part of the *bla* gene plus the FRT site from pSG76-A to create pLOI2216. Digestion, Klenow treatment and self-ligation [pLOI2217] abolished the *HindIII* site in pLOI2216. The restriction sites from *EcoRI* to *PstI* were removed by double-digestion with these enzymes, *Klenow* treatment and self-ligation [pLOI2218]. Then a new *EcoRI* site was created by linker addition into the 30 previously added *ClaI* site, giving rise to a pUC18- plasmid derivative carrying one FRT site (called FRT1 in Table 1) but no MCS [pLOI2219].

- 26 -

To introduce a new MCS into pLOI2219, the *Sph*I site from pUC18 was replaced with an *Asc*I site by linker addition [pLOI2400], and the *Kpn*I site was replaced with a *Bgl*II linker [pLOI2401]. The modified pUC18-based MCS was then introduced into pLOI2219 as a 965 bp *Scal/Eco*RI fragment originating from pLOI2220. By linker
5 addition, an *Asc*I site took the place of *Eco*RI site [pLOI2221], and a new *Eco*RI site replaced the *Hind*III site [pLOI2222].

The resultant plasmid was sequenced in both direction and used as a source of DNA carrying the modified MCS and the FRT1 site.

10 *Construction of pLOI2223*

To create the conditional replication vector carrying two FRT sites and the ampicillin-resistance gene, pLOI2222 was double-digested with *Eco*RI and *Scal* and a 157 bp fragment was ligated to pSG76-A (R6K- γ ori, Ap^R) digested with the same enzymes. Transformation was done in *E. coli* S17-1, since this strain provides the λ *pir*
15 protein involved in the replication of the R6K- γ ori and selection was done in the presence of 50 μ g/ml ampicillin. The resultant plasmid was denominated pLOI2223.

Construction of pLOI2224

To create the conditional replication vector carrying two FRT sites and the kanamycin-resistance gene, pLOI2222 was double-digested with *Eco*RI and *Scal* and a 157 bp fragment was ligated to pSG76-K (R6K- γ ori, Km^R) digested with the same enzymes. Transformation was done in *E. coli* S17-1, since this strain provides the λ *pir* protein involved in the replication of the R6K- γ ori and selection was done in the presence of 50 μ g/ml kanamycin. DNA from several colonies was prepared and
25 analyzed by digestion with *Asc*I and *Bgl*II, separately, to confirm the presence of the modified MCS. The resultant plasmid was denominated pLOI2224.

Construction of pLOI2225

To create the conditional replication vector carrying two FRT sites and the chloramphenicol-resistance gene, pLOI2222 was double-digested with *Eco*RI and *Scal* and a 157 bp fragment was ligated to pSG76-C (R6K- γ ori, Cm^R) digested with the same enzymes. Transformation was performed as described above and selection was done in
30

- 27 -

the presence of 50 µg/ml chloramphenicol. The resultant plasmid was termed pLOI2225.

Construction of pLOI2226

- 5 To create the thermosensitive replication vector carrying two FRT sites and the ampicillin-resistance gene, pLOI2222 was double-digested with *Eco*RI and *Scal* and a 157 bp fragment was ligated to pST76-A (pSC101 *ori*, Ap^R) digested with the same enzymes. For plasmids carrying the thermosensitive replicon (pSC101), *E. coli* S17-1 transformed cells were kept at 30°C and cells were selected with 40 µg/ml ampicillin.
- 10 The resultant plasmid was termed pLOI2226.

Construction of pLOI2227

- To create the thermosensitive replication vector carrying two FRT sites and the kanamycin-resistance gene, pLOI2222 was double-digested with *Eco*RI and *Scal* and a 157 bp fragment was ligated to pST76-K (pSC101 *ori*, Km^R) digested with the same enzymes and transformation was performed as described above and selection was done in the presence of 40µg/ml of kanamycin. The resultant plasmid was termed pLOI2227.

Construction of pLOI2228

- 20 To create the thermosensitive replication vector carrying two FRT sites and the chloramphenicol-resistance gene, pLOI2222 was double-digested with *Eco*RI and *Scal* and a 157 bp fragment was ligated to pST76-C (pSC101 *ori*, Cm^R) digested with the same enzymes and transformation was performed as described above and selection was done in the presence of 40µg/ml of chloramphenicol. The resultant plasmid was termed
25 pLOI2228.

Construction of PLOI2231

- The construct pLOI2231 (SEQ ID NO: 7) is a derivative of pLOI2224 where *adhE* (guide) and the PET (passenger) genes have been cloned. The PET passenger
30 genes comprise *adhB* and *pdc*. To obtain this construction the *adhE* gene was first cloned into pCR2.1-TOPO vector as indicated by manufacturer. Briefly, the *adhE* gene was PCR amplified using a pair of Genosys ORFmer primers:

- 28 -

5 'TTGCTCTTCCATGGCTGTTACTAATGTCGCTGAA3' forward primer;
5 'TTGCTCTTCGTTAACGGATTTTCGCTTTCT3' reverse primer. The PCR product of *adhE* includes the initiation codon ATG and the termination codon TAA. A fresh PCR sample of *adhE* was ligated to the pCR2.1-TOPO vector. This plasmid was
5 designated as pLOI2408. A subcloning step was performed to transfer *adhE* from pLOI2408 to pLOI2403. Plasmid pLOI2408 was digested with restriction enzyme *Eco*RI. The product of this digestion allows to obtain the whole PCR product previously cloned. The *adhE* fragment was separated by gel electrophoresis and isolated from the gel using the QIAGEN kit.

10 The isolated *adhE* fragment was visualized by UV light after ethidium bromide staining and gel electrophoresis. This fragment was ligated to plasmid pLOI2403. The construct pLOI2403 was previously digested with restriction enzyme *Eco*RI. The *adhE* gene fragment and pLOI2403 were ligated using the rapid DNA ligation kit as directed by manufacturer. The plasmid obtained after the ligation between the *adhE Eco*RI
15 digested fragment and pLOI2403 *Eco*RI digested plasmid was designated pLOI2413. In this plasmid, the 3' end of the *adhE* gene is orientated to be close to the *Bam*HI site of pLOI2403 as revealed by sequence analysis.

In order to ligate the PET genes (passenger) to *adhE* gene (guide) a ligation was performed between a PET fragment and the plasmid pLOI2413. PET genes were
20 isolated from plasmid pLOI510 as a *Bam*HI fragment of about 4.4 Kbp. The PET *Eco*RI digested fragment was isolated from a gel after electrophoresis of plasmid pLOI510 *Bam*HI digested with the QIAGEN kit. The PET DNA fragment was visualized by ethidium bromide after gel electrophoresis. A ligation of the isolated PET fragment with plasmid pLOI2413 previously digested with restriction enzyme *Bam*HI was performed.
25 The plasmid obtained was designated pLOI2230.

To obtain an integration vector with guide and the passenger fragments as shown in Fig. 3, the plasmid pLOI2224 was ligated to a fragment containing *adhE* and PET genes isolated from plasmid pLOI2230. The resultant plasmid termed pLOI2230 was digested with the restriction enzyme *Asc*I and the digestion product was cloned into
30 pLOI2224 previously digested with enzyme *Asc*I. The plasmid obtained from this ligation was designated pLOI2231 and the orientation of the various genetic elements of the plasmid are shown in Fig. 3

Targeting Strategy

Recombinase-based integration systems offer the opportunity to effect precise DNA deletions *in vivo* (Cherepanov, P. P., *et al.* *Gene* 158: 9-14 (1995); Posfai, G. *et al.*, 5 *Nucleic Acids Res.* 22: 2392-2398 (1994)) and *in vitro* (Cox, M. M., *Proc. Natl. Acad. Sci. USA* 80: 4223-4227 (1983); Hasan, N., *et al.*, *Gene* 150: 51-56 (1994), Wild, J., Z. *et al.*, *Gene* 179:181-188 (1996)). The present invention encompasses plasmids as described by Posfai *et al.* (*J. Bacteriol.* 179: 4426-4428 (1997)) that were further modified to incorporate homologous DNA as a guide sequence, two FRT sites in the 10 same orientation flanking an antibiotic marker and a multi cloning site (MCS) (Fig. 1).

A full set of plasmids (ampicillin, chloramphenicol, and kanamycin resistance) was made for each conditional replicon (Fig. 1; pSC101 and R6K- γ). Since both conditional replicons are present at low copy numbers, an additional high copy vector was developed to facilitate constructions by adding an *AscI* site on either side of the 15 MCS in pLITMUS38 (New England Biolabs, Beverly, MA) to produce pLOI2403 (Fig. 1).

A general procedure for chromosomal integration of DNA using the new vectors with conditional replicons is presented in Fig. 2. The integration of heterologous passenger DNA carrying desired functions can be targeted to any specific chromosomal 20 site by an adjacent fragment of homologous DNA (guide) during growth at 30°C (pSC101), or directly selected at 37-42°C (pSC101 or R6K- γ). With a single cross-over event, the entire plasmid is incorporated into the chromosome (If needed, pSC101-based integration vectors can be eliminated by overnight growth and plating at elevated temperatures.).

25 After integration, recombinants were transformed with pFT-A containing the yeast FLP gene under control of the tetracycline promoter and grown under permissive conditions (30°C, pSC101). During growth with chlortetracycline, FLP recombinase was induced which in turn resulted in the excision of the DNA bracketed by the concurrently-facing FRT sites (selectable marker and replicon) from the chromosome as 30 shown in Fig. 3. After growth at 37-42°C to eliminate plasmid pFT-A encoding the recombinase, only the passenger gene(s), a single FRT, and the homologous guide fragment remained in the chromosome as described in further detail below.

- 30 -

EXAMPLE II

A Method for Engineering an Ethanologenic Host Cell by Chromosomal Integration of Heterologous DNA

5

In this example, methods for introducing heterologous DNA to engineer an ethanologenic organism are described.

To illustrate the utility of the foregoing vectors, heterologous genes were targeted into the genome of two different test microorganism in order to generate a 10 genetically engineered organism with a desired phenotype (*i.e.*, ability to produce ethanol). In particular, derivatives of *E. coli* B (strain SE2272, *Δfrd*) and *E. coli* K-12 (SE2275, *Δfrd*) where constructed in which three heterologous genes were integrated immediately behind the endogenous *adhE* gene in the chromosome. The guide and passenger DNA were cloned into pLOI2403, a high copy plasmid vector described 15 above. For this construction, the promoterless *adhE* coding region (guide) was amplified with Genosys ORFmer primers for the coding region (forward, 5' TTGCTCTTCCATGGCTGTTACTAATGTCGCTGAA3'; reverse, 5' TTGCTCTCGTTAACGGATTTTCGCTTTCT3') and cloned into pCR2.1-TOPO to produce pLOI2408 (see for example Fig. 3). After *Eco*RI digestion, the 2.6 Kbp *adhE* 20 region from pLOI2408 was moved into the corresponding site in pLOI2403 to produce pLOI2413. The *Bam*HI site immediately downstream from the 3' end of the *adhE* coding region was used to insert a 4.6 Kbp *Bam*HI fragment from pLOI510 (Ohta, K. *et al.*, *Appl. Environ. Microbiol.* 57: 893-900 (1991)) containing three genes (passenger): a promoterless *Zymomonas mobilis* *pdc* without transcriptional terminator and a 25 promoterless *Z. mobilis* *adhB* with a transcriptional terminator, followed by a complete *cat* operon with promoter and terminator. In the resulting plasmid (pLOI2230), transcription of the heterologous genes was oriented concurrently with *adhE*. All constructs containing the *Z. mobilis* genes were grown in LB supplemented with glucose (20 g/L for plates; 50 g/L for broth).

30 The 7.2 Kbp *Asc*I fragment from pLOI2230 (high copy vector) containing *adhE*, the artificial operon *pdc adhB*, and *cat* was ligated into the low copy integration vector pLOI2224 which contains a R6K-γ replicon (*λpir*-dependent) and transformed into the

- 31 -

permissive host S17-1 (de Lorenzo, V., *et al.*, *J. Bacteriol.* 172: 6568-6572 (1990)) with selection for kanamycin and chloramphenicol. The resulting clone containing pLOI2231 was used for large-scale plasmid isolation (500 ml) by the alkaline lysis procedure (Sambrook, J., *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor 5 *Laboratory Press, Cold Spring Harbor, NY*(1989)).

Approximately 500 ng of pLOI2231 DNA was used for electroporation of SE2272 and SE2275. Both are non-permissive hosts. Recombinants were readily obtained by selection for either kanamycin (vector) or chloramphenicol (passenger) resistance. Up to 2 h was allowed for expression of the resistance gene prior to 10 spreading on plates for selection. Approximately 1000 recombinants per 1 µg DNA (electroporation) were recovered using *E. coli* K-12 SE2275, 5-fold higher than that obtained with the *E. coli* B SE2272. Thirty recombinants from each host were screened for the functional expression of alcohol dehydrogenase on indicator plates (Conway, T.C., *et al.*, *J. Bacteriol.* 169:2591-2597 (1987)). Based on the rate and intensity of 15 color development, these recombinants expressed higher levels of ADH activity than the respective unmodified SE2272 or SE2275, or S17-1(pLOI2231) harboring promoterless *pdc* and *adhB* genes. Unlike the control strains, these recombinants also exhibited a colonial phenotype (large raised colonies on LB containing glucose) that is typical for ethanologenic *E. coli* (Ingram, L.O., *et al.*, *Environ. Microbiol.* 53: 2420-2425 (1987)). 20 Small-scale DNA preparations (7 recombinants per host) were tested for the presence of pLOI2231. None contained plasmids as tested by gel filtration or based on transformation experiments using S17-1 as the host. These recombinants were presumed to contain chromosomally integrated genes. One clone from each parent, strains FM7 (*E. coli* B SE2272) and FM19 (*E. coli* K-12 SE2275), was selected for 25 further study.

Accordingly, strains FM7 and FM19 were transformed with the helper plasmid (pFT-A) carrying the FLP gene (Posfai, G., *et al.*, *J. Bacteriol.* 179: 4426-4428 (1997)) and incubated at 30°C with selection for ampicillin resistance. A mixture of colonies was used to inoculate a broth culture for induction of FLP with autoclaved 30 chlortetracycline (20 µg/ml). After 6 h incubation at 30°C, the culture was diluted 1:1000 in LB containing glucose and incubated at 42°C for 16 h to eliminate the helper plasmid. After streaking on solid medium, isolated colonies were screened for the

- 32 -

absence of antibiotic markers. Approximately 80% of the colonies were ampicillin and kanamycin sensitive and retained only chloramphenicol resistance and the ethanologenic traits (passenger genes inserted into the MCS). Loss of ampicillin resistance indicated that the helper plasmid had been successfully eliminated while loss of kanamycin
5 resistance confirmed the FLP-recombinase-dependent deletion of the vector. These new derivatives of FM7 and FM19 were designated FM18 and FM20, respectively.

PCR was used to verify the integration events in both FM18 and FM20. Two new sets of primers were designed to amplify the *adhE* gene including the unique junctions predicted for *pdc* (Fig. 3, primers 3 and 4) and *cat* (Fig. 3, primers 5 and 6) as
10 a result of integration and FLP-mediated deletion. Forward primer 3 hybridizes to the promoter region of *adhE* while reverse primer 4 hybridizes to the N-terminal coding region of *pdc*. Forward primer 5 hybridizes to the C-terminal coding region of *cat* gene and reverse primer 6 hybridizes to the 3' untranslated portion of the *adhE*. Note that the primers used to clone *adhE*, forward primer 1 and reverse primer 2, hybridize to the N-
15 terminal and C-terminal coding regions of *adhE* and are inside of the regions encoded by forward primer 3 and reverse primer 6. All primer sets (SE2272 template: 1+2, 3+6; FM18 template: 3+4, 5+6) generated products of the expected sizes (Fig. 4A). Identical results were also obtained using FM20 DNA as the template. As shown in Fig. 3, and arrows with numbers 1, 2, 3, 4, 5, and 6 represent primers used to amplify the
20 corresponding regions. The sequences for these primers are as follows: Primer 1.

5' TTGCTCTCCATGGCTGTTACTAATGTCGCTGAA3'; Primer 2.

5' TTGCTCTCGTTAAGCGGATTTTCGCTTTTCT3'; Primer 3.

5' GTGAGTGTGAGCGCGGAGT3'; Primer 4. 5' TGGCACGAGCATAACCTTC3'; Primer 5.

5' CAGTACTGCGATGAGTGGCA3'; Primer 6. 5' GTTGCCAGACAGCGCTACT3'.

25 The *adhE* gene contains a single central *Bst*EII site which does not occur elsewhere in the PCR products. This site was used to verify the identity of the PCR fragments. As shown in Fig. 4B, all PCR products were cut once to produce fragments containing the N-terminal and C-terminal regions of *adhE*. Fragments from the *adhE* coding region alone (primers 1+2) were smaller (N-terminal fragment = 1,226 bp; C-
30 terminal fragment = 1,470 bp) than fragments which included parts of the native *adhE* promoter (primers 3+4 and 3+6; N-terminal fragment = 1,325 bp) or *adhE* terminator (primers 3+6, 5+6; C-terminal fragment = 1,489 bp). The fragment which included part

- 33 -

of *pdc* (primers 3+4) was the largest C-terminal fragment (1,783 bp). The fragment which included part of *cat* (primers 5+6) was the largest N-terminal fragment (1,988 bp).

In Fig. 4, panel A shows full length PCR products of *adhE* junctions. Positions and sequences of primers are provided in Fig. 3. The lanes in panel A of Fig. 4 represent: 1., *Hind*III digest of phage λ DNA; 2., *adhE* coding region (2,696 bp) of SE2272 amplified using forward primer 1 and reverse primer 2; 3., *adhE* promoter and 3' untranslated sequence (2,814 bp) of SE2272 amplified using forward primer 3 and reverse primer 4; 4., The *adhE* and *pdc* junction (3,108 bp) of FM18 amplified using forward primer 3 and reverse primer 4; 5., The *cat* junction and *adhE* (3,477 bp) of FM18 amplified using forward primer 5 and reverse primer 6. The amplification conditions for primer pairs 1 + 2 and 3 + 6 (25 cycles) were: 45 sec at 94°C, 45 sec at 60°C, 60 sec at 72°C. The amplification conditions for primer pairs 3 + 4 and 5 + 6 (30 cycles) were: 45 sec at 94°C, 60 sec at 60°C. DNA was held at 94°C for 3 min prior to the first cycle. The elongation time was increased to 10 min during the final cycle.

In Fig. 4, panel B shows a *Bst*EII digestion of the above-described PCR products. A single, central *Bst*EII site was used to cleave PCR products containing *adhE* into a N-terminal and C-terminal fragments. The lanes in panel B of Fig. 4 show the following: Lane 1 contains a DNA standard (descending: 3.0, 2.0, 1.5, 1.2, 1.0 Kbp); Lanes 2 through 5 contain *Bst*EII digested PCR products of fragments described in 4A, respectively.

The expression of *adhE* is regulated by a number of factors in *E. coli* including *cra*, *adhR*, and the abundance of NADH (Leonardo, M.R., et al., *J. Bacteriol.* 175: 870-878 (1993); Leonardo, M.R., et al., *J. Bacteriol.* 178: 6013-6018 (1996); Mikulskis, A., et al., *J. Bacteriol.* 179: 7129-7134 (1997)). Both message levels and activity are approximately 10-fold higher during anaerobic growth with glucose than during aerobic growth. Since the *Z. mobilis* genes are integrated behind the *adhE* coding region to form an operon fusion, expression of *pdc* should also increase in response to anaerobiosis. Strains FM18 and FM20 were grown in Luria broth containing 50 g glucose/liter under aerobic and anaerobic conditions. PDC activities were determined in heat-treated preparations to eliminate lactate dehydrogenase and other confounding activities (Conway, T., et al., *J. Bacteriol.* 169: 949-954 (1987)). Under anaerobic conditions, PDC activities in FM18 and FM20 were 0.254 U/mg protein and 0.185 U/mg protein,

- 34 -

respectively, approximately 4-fold higher than the activities observed in cells grown under aerobic conditions.

The above results demonstrate that the new integration vectors can be used to place promoterless genes under the control of a chromosomal promoter in a site specific fashion. Moreover, after the integration event, unnecessary nucleic acid encoding a replicon and selectable marker was removed *in vivo* using a recombinase site specific recombinase. This approach avoids potential problems of lethality or mutation due to unregulated expression in plasmids during construction and integration. These vectors can also be used to replace promoters in chromosomal genes. Additional unique restriction sites are available for the insertion of genes which can be temporarily expressed after integration and subsequently deleted with the replicon and selectable marker. This option, the temporary introduction of new genes, may be useful to test new traits in an isogenic background.

Although the vectors described must be propagated in *E. coli*, they are potentially useful with other organisms. The FLP recombinase is extremely efficient (Wild, J., Z., *et al.*, *Gene* 179:181-188 (1996)) and could be produced intracellularly as a transient expression product after transformation or electroporation of pFT-A.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims. Moreover, any number of genetic constructs, host cells, and methods described in United States Patent Nos. 5,821,093; 5,482,846; 5,424,202; 5,028,539; 5,000,000; 5,162,516; and U.S. patent application serial No. 60/136,376 may be employed in carrying out the present invention and are hereby incorporated by reference.

Claims

1. A method for integrating a nucleic acid construct into the genome of a host cell, said method comprising,
 - 5 contacting said cell with a nucleic acid construct under conditions such that said nucleic acid construct is integrated by said cell, wherein said nucleic acid construct comprises a passenger sequence and a marker sequence, wherein said marker sequence is flanked by a first and second recombining site wherein said recombining sites are oriented in the same direction.
- 10 2. The method of claim 1, wherein said construct further comprises an origin of replication between said first and second recombining sites.
- 15 3. The method of claim 2, wherein said origin of replication is, or is derived from, an origin of replication selected from the group consisting of pSC101ori, R6K- γ ori, colE1, and oriEV.
- 20 4. The method of claim 1, wherein said nucleic acid further comprises a sequence selected from the group consisting of a promoter, a restriction site, an intron, an IRES element, and a polyadenylation site.
5. The method of claim 1, wherein said nucleic acid further comprises a guide sequence thereby resulting in site-specific integration of said nucleic acid construct.
- 25 6. The method of claim 1, wherein said guide sequence, comprises genomic sequence derived from a replicating genome.
7. The method of claim 1, wherein said cell is provided with a site-specific recombinase that results in recombination between said first and second recombining sites, wherein said sequence flanked by said recombining sites is excised.

- 36 -

8. The method of claim 7, wherein said recombinase is selected from the group consisting of FLP, Xer, Int, and Cre.

9. The method of claim 7, wherein said recombinase is FLP recombinase.

5

10. The method of claim 1, wherein said recombining site is selected from the group consisting of FRT, dif, att, and loxP.

11. The method of claim 1, wherein said recombining site is FRT.

10

12. The method of claim 7, wherein following excision of said sequence, the method of claim 1 is repeated on said cell.

15

13. The method of claim 1, wherein said construct further comprises a promoter 5' to the passenger sequence.

14. The method of claim 1, wherein said marker sequence is selected from the group consisting of a antibiotic resistance gene and a non-antibiotic resistance gene.

20

15. The method of claim 14, wherein said antibiotic resistance gene is selected from the group consisting of the gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, and chloramphenicol resistance gene.

25

16. The method of claim 11, wherein said non-antibiotic resistance gene is selected from the group consisting of an auxotrophic gene, a metal ion resistance gene, a *trp* gene, and a gene encoding a green fluorescent protein.

30

17. The method of claim 1, wherein said passenger sequence encodes at least one gene.

18. The method of claim 17, wherein said passenger sequence encodes a gene involved in ethanogenesis.

19. The method of claim 18, wherein said gene involved in ethanogenesis is 5 derived from an organism selected from the group consisting of prokaryotes and eukaryotes.

20. The method of claim 18, wherein said gene is selected from the group consisting of *adh* and *pdc*.

10

21. The method of claim 1, wherein said host cell is a bacterial cell.

22. The method of claim 1, wherein said host cell is a Gram-negative bacterial cell.

15 23. The method of claim 22, wherein said host cell is a facultatively anaerobic bacterial cell.

24. The method of claim 23, wherein said host cell is selected from the family Enterobacteriaceae.

20

25. The method of claim 24 wherein said host is selected from the group consisting of *Escherichia* and *Klebsiella*.

26. The method of claim 24, wherein said host is *E. coli* B or *E. coli* K12.

25

27. The method of claim 1, wherein said host cell is a recombinant bacterial cell.

28. The method of claim 1, wherein said nucleic acid construct is, or is derived from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225,

30 pLOI2226, pLOI2227, pLOI2228, pLOI2231, and pLOI2403.

29. A method for producing a recombinant ethanologenic cell, said method comprising,

contacting said cell with a nucleic acid construct under conditions in which integration of said nucleic acid construct occurs forming a recombinant ethanologenic cell, wherein said nucleic acid construct comprises a passenger sequence, wherein said passenger sequence comprises an ethanologenic gene, and a marker sequence, wherein said marker sequence is flanked by a first and second recombining site.

30. The method of claim 29, wherein said construct further comprises an origin of replication between said first and second recombining site.

31. The method of claim 30, wherein said origin of replication is, or is derived from, an origin of replication selected from the group consisting of pSC101ori, R6K- γ ori, colE1ori, and oriEV.

15
32. The method of claim 29, wherein said cell is provided with a site-specific recombinase that results in recombination between said first and second recombining sites wherein said sequence flanked by said recombining sites is excised.

20 33. The method of claim 29, wherein said recombinase is selected from the group consisting of FLP, Xer, Int, and Cre.

34. The method of claim 33, wherein said recombinase is FLP recombinase.

25 35. The method of claim 29, wherein said recombining site is selected from the group consisting of FRT, dif, att, and loxP.

36. The method of claim 35, wherein said recombining site is FRT.

30 37. The method of claim 29, wherein said ethanologenic gene is selected from the group consisting of *adh* and *pdc*.

- 39 -

38. The method of claim 29 where said passenger sequence encodes *pdc*, *adhB*, and *cat*.
39. The method of claim 29, wherein said nucleic acid further comprises a guide sequence thereby resulting in site-specific integration of said nucleic acid construct.
5
40. The method of claim 39, wherein said guide sequence is derived from a replicating genome.
- 10 41. The method of claim 29, wherein said marker sequence is selected from the group consisting of a antibiotic resistance gene or a non-antibiotic resistance gene.
42. The method of claim 41, wherein said antibiotic resistance gene is selected from the group consisting of gentamycin resistance gene, zeocin resistance gene, kanamycin
15 resistance gene, ampicillin resistance gene, tetracycline resistance gene, and chloramphenicol resistance gene.
43. The method of claim 41, wherein said non-antibiotic resistance gene is selected from the group consisting of an auxotrophic gene, a metal ion resistance gene, a *trp* gene, and a gene encoding a green fluorescent protein.
20
44. The method of claim 41, where said marker sequence comprises the kanamycin resistance gene.
- 25 45. A recombinant ethanologenic cell produced according to the method of claim 29.
46. A recombinant host cell comprising,
a nucleic acid construct that comprises a passenger sequence and a marker sequence, wherein said marker sequence is flanked by a first and second
30 recombining site wherein said recombining sites are oriented in the same direction.

- 40 -

47. The host cell according to 46, wherein said construct further comprises an origin of replication between said first and second recombining sites.

48. The host cell according to 46, wherein said origin of replication is, or is derived
5 from, an origin of replication selected from the group consisting of pSC101ori, R6K- γ ori, colE1, and oriEV.

49. The host cell according to claim 46, wherein said nucleic acid construct is integrated.

10

50. The host cell according to claim 46, wherein said nucleic acid further comprises a guide sequence thereby resulting in site-specific integration of said nucleic acid construct.

15

51. The host cell according to claim 46, wherein said guide sequence, comprises genomic sequence derived from a replicating genome.

52.

The host cell according to claim 46, wherein said cell is provided with a site-specific recombinase that results in recombination between said first and second
20 recombining sites, wherein said sequence flanked by said recombining sites is excised.

53. The host cell according to claim 46, wherein said recombinase is selected from the group consisting of FLP, Xer, Int, and Cre.

25

54. The host cell according to claim 46, wherein said recombinase is FLP recombinase.

55. The host cell according to claim 46, wherein said recombining site is selected from the group consisting of FRT, dif, att, and loxP.

30

56. The host cell according to claim 46, wherein said recombining site is FRT.

- 41 -

57. The host cell according to claim 46, wherein said marker sequence is selected from the group consisting of a antibiotic resistance gene or a non-antibiotic resistance gene.

5 58. The host cell according to claim 57, wherein said antibiotic resistance gene is selected from the group consisting of a gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, and chloramphenicol resistance gene.

10 59. The host cell according to claim 57, wherein said non-antibiotic resistance gene is selected from the group consisting of an auxotrophic gene, a metal ion resistance gene, a *trp* gene, and a gene encoding a green fluorescent protein.

15 60. The host cell according to claim 46, wherein said passenger sequence encodes a gene involved in ethanogenesis.

61. The host cell according to claim 46, wherein said passenger sequence is selected from the group comprising *adh*, *pdc*, and *cat*.

20 62. The host cell according to claim 46, wherein said host cell is a bacterial cell.

63. The host cell according to claim 46, wherein said host cell is a Gram-negative bacterial cell.

25 64. The host cell according to claim 46, wherein said host cell is a facultatively anaerobic bacterial cell.

65. The host cell according to claim 46, wherein said host cell is selected from the family Enterobacteriaceae.

30

66. The host cell according to claim 46, wherein said host is selected from the group consisting of *Escherichia* and *Klebsiella*.

- 42 -

67. The host cell according to claim 46, wherein said host is *E. coli* B or *E. coli* K12.

68. The host cell according to claim 46, wherein said host cell is a recombinant bacterial cell.

5

69. The host cell according to claim 46, wherein said nucleic acid construct is, or is derived from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225, pLOI2226, pLOI2227, pLOI2228, pLOI2231, and pLOI2403.

10 70. The host cell according to claim 46, wherein said cell is ethanologenic.

71. A method for producing ethanol comprising,
providing a recombinant ethanologenic cell having a nucleic acid construct that comprises a passenger sequence and a marker sequence, wherein said 15 marker sequence is flanked by a first and second recombining site, and
contacting said cell with a substrate which can be fermented into ethanol, wherein expression of said passenger sequence results in the production of ethanol.

20 72. The method of claim 71, wherein said construct further comprises an origin of replication between said first and second recombining sites.

73. The method of claim 72, wherein said origin of replication is, or is derived from, an origin of replication selected from the group consisting of pSC101ori, R6K- γ ori, 25 colE1, and oriEV.

74. The method according to claim 71, wherein said nucleic acid construct is integrated.

30 75. The method according to claim 71, wherein said nucleic acid construct further comprises a guide sequence thereby resulting in site-specific integration of said nucleic acid construct.

- 43 -

76. The method according to claim 75, wherein said guide sequence, comprises genomic sequence derived from a replicating genome.

77. The method according to claim 71, wherein said cell is provided with a site-specific recombinase that results in recombination between said first and second recombining site, wherein said sequence flanked by said recombining sites is excised.

78. The method according to claim 71, wherein said recombinase is selected from the group consisting of FLP, Xer, Int, and Cre.

10

79. The method according to claim 71, wherein said recombinase is FLP recombinase.

15

80. The method according to claim 71, wherein said recombining site is selected from the group consisting of FRT, dif, att, and loxP.

81. The method according to claim 71, wherein said recombining site is FRT.

82. The method according to claim 71, wherein said marker sequence is selected from the group consisting of a antibiotic resistance gene or a non-antibiotic resistance gene.

83. The method of claim 82, wherein said antibiotic marker is selected from the group consisting of a gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, and chloramphenicol resistance gene.

84. The method of claim 82, wherein said non-antibiotic resistance gene is selected from the group consisting of an auxotrophic gene, a metal ion resistance gene, a *trp* gene, and a gene encoding a green fluorescent protein.

30

- 44 -

85. The method according to claim 71, wherein said passenger sequence encodes a gene involved in ethanogenesis.

86. The method according to claim 71, wherein said passenger sequence is selected
5 from the group comprising *adh* and *pdc*.

87. The method according to claim 71, wherein said passenger sequence encodes
adhB, *pdc*, and *cat*.

10 88. The method according to claim 71, wherein said host cell is a bacterial cell.

89. The method according to claim 71, wherein said host cell is a Gram-negative
bacterial cell.

15 90. The method according to claim 71, wherein said host cell is a facultatively
anaerobic bacterial cell.

91. The method according to claim 71, wherein said host cell is selected from the
family Enterobacteriaceae.

20

92. The method according to claim 71, wherein said host is selected from the group
consisting of *Escherichia* and *Klebsiella*.

93. The method according to claim 71, wherein said host is *E. coli* B or *E. coli* K12.

25

94. The method according to claim 71, wherein said host cell is a recombinant
bacterial cell.

95. The method according to claim 71, wherein said nucleic acid construct is, or is
30 derived from, a plasmid selected from the group consisting of pLOI2223, pLOI2224,
pLOI2225, pLOI2226, pLOI2227, pLOI2228, pLOI2231, and pLOI2403.

- 45 -

96. A nucleic acid construct comprising,
a passenger sequence, and
a marker sequence, wherein said marker sequence is flanked by a first
and second recombining site.

5

97. The construct according to claim 96, wherein said construct further comprises an
origin of replication between said first and second recombining sites.

98. The construct according to claim 97, wherein said origin of replication is, or is
10 derived from, an origin of replication selected from the group consisting of pSC101ori,
R6K- γ ori, colE1, and oriEV.

99. The construct according to claim 84, wherein said construct further comprises a
guide sequence.

15

100. The construct according to claim 99, wherein said guide sequence comprising a
sequence derived from a replicating genome.

101. The construct according to claim 99, wherein said guide sequence is derived
20 from a bacterial cell.

102. The construct according to claim 96, wherein said construct further comprises a
at least one unique restriction enzyme site.

25 103. The construct according to claim 96, wherein said passenger sequence comprises
an ethanologenic gene.

104. The construct according to claim 96, wherein said passenger sequence comprises
a gene selected from the group consisting of *adh* and *pdc*.

30

105. The construct according to claim 96, wherein said passenger sequence comprises
adhB, *pdc*, and *cat*.

- 46 -

106. The construct according to claim 96, wherein said recombining site is selected from the group consisting of FRT, dif, att, and loxP.

107. The construct according to claim 96, wherein said recombining site is FRT.

5

108. The construct according to claim 96, wherein said construct is exposed to a site-specific recombinase such that a recombination between said first and second recombining site occurs, wherein said sequence flanked by said recombining sites is excised.

10

109. The construct according to claim 108, wherein said recombinase is selected from the group consisting of FLP, Xer, Int, and Cre.

110. The construct according to claim 109, wherein said recombinase is FLP.

15

111. The construct according to claim 96, wherein said passenger sequence comprises a sequence selected from the group consisting of a heterologous promoter and a prokaryotic termination sequence.

20 112. The construct according to claim 96, wherein said marker sequence is selected from the group consisting of a antibiotic resistance gene or a non-antibiotic resistance gene.

25 113. The construct according to claim 112, wherein said antibiotic resistance gene is selected from the group consisting of a gentamycin resistance gene, zeocin resistance gene, kanamycin resistance gene, ampicillin resistance gene, tetracycline resistance gene, and chloramphenicol resistance gene.

30 114. The construct according to claim 112, wherein said non-antibiotic resistance gene is selected from the group consisting of an auxotrophic resistance gene, a metal ion resistance gene, a *trp* gene, and a gene encoding a green fluorescent protein.

- 47 -

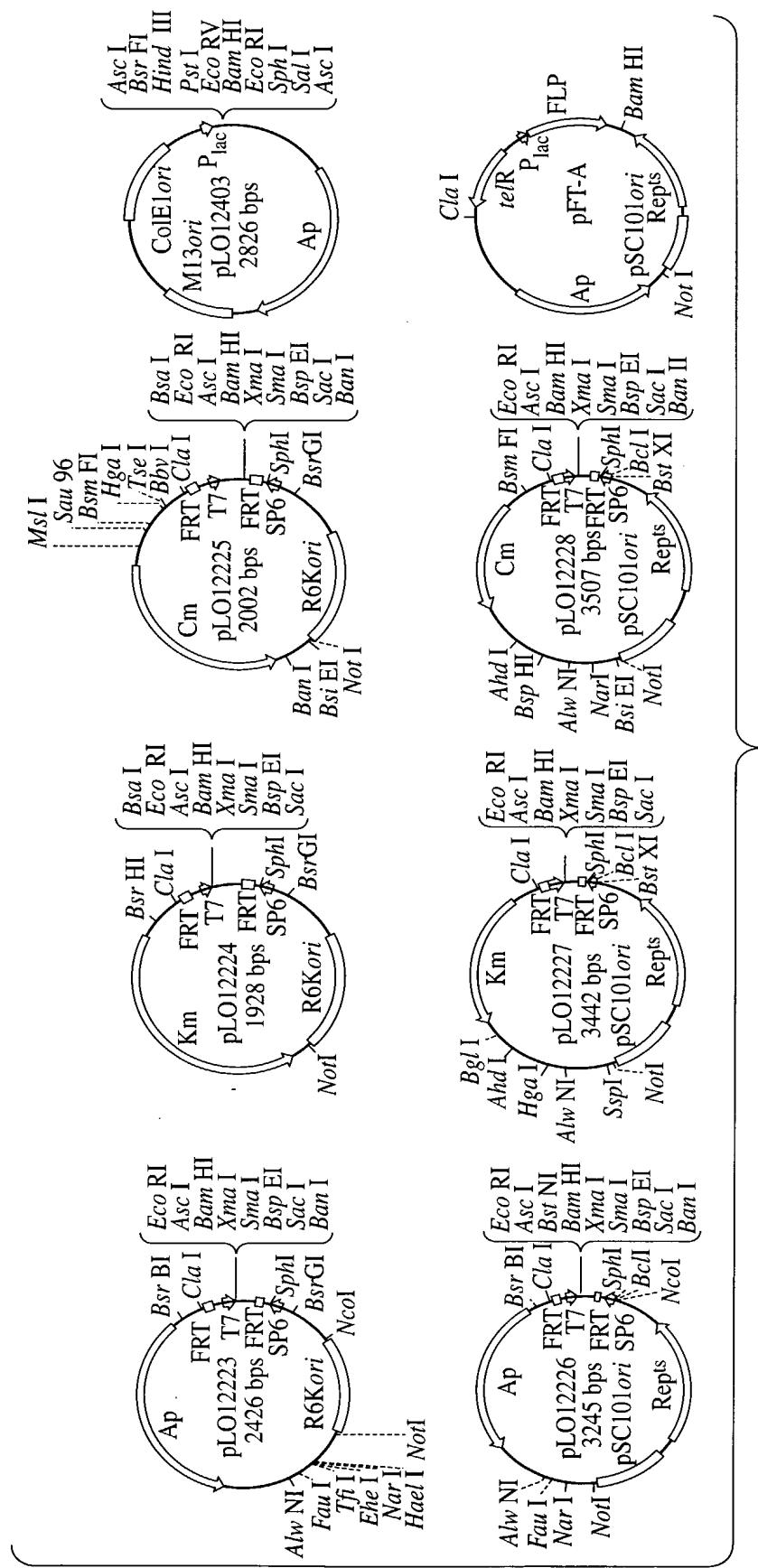
115. The construct according to claim 96, wherein said marker sequence comprises the kanamycin resistance gene.

116. The construct according to claim 96, wherein said construct is, or is derived
5 from, a plasmid selected from the group consisting of pLOI2223, pLOI2224, pLOI2225,
pLOI2226, pLOI2227, pLOI2228, pLOI2231, and pLOI2403.

117. A kit comprising at least one nucleic acid construct according to claim 96 and
instructions for use.

10

1/4

**Fig. 1**

2/4

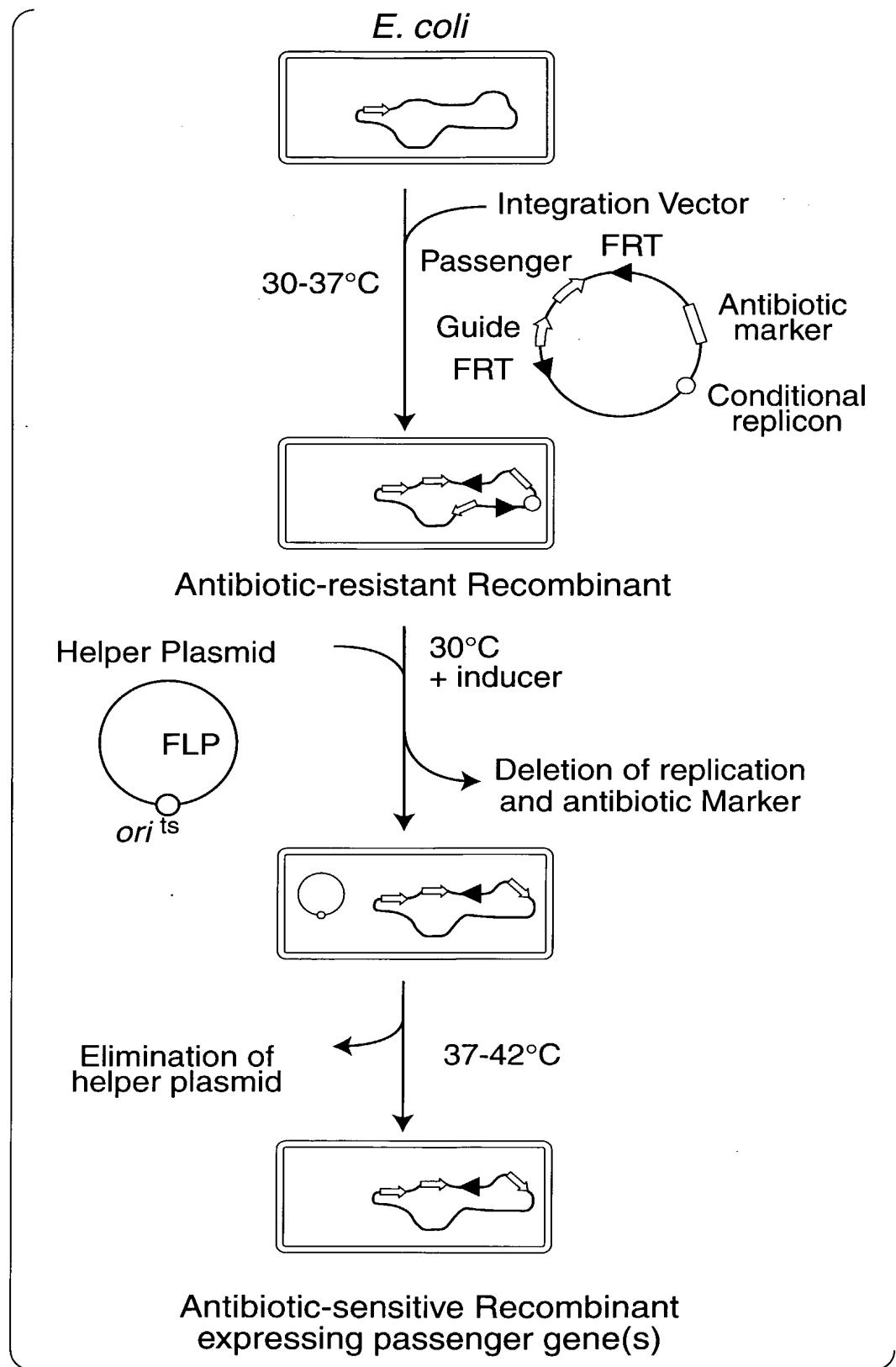


Fig. 2

3/4

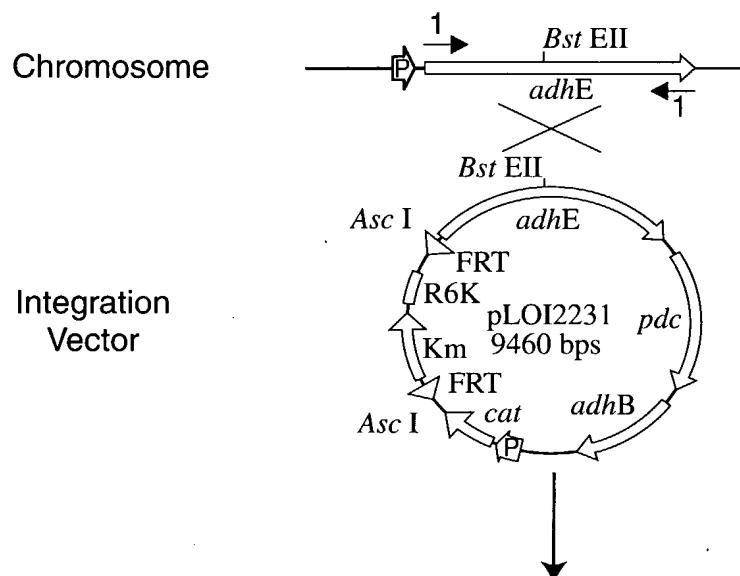


Fig. 3A

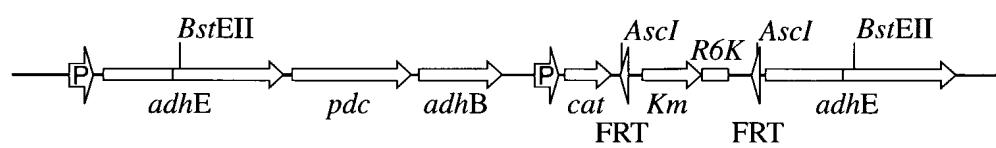


Fig. 3B

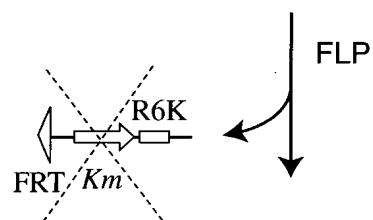
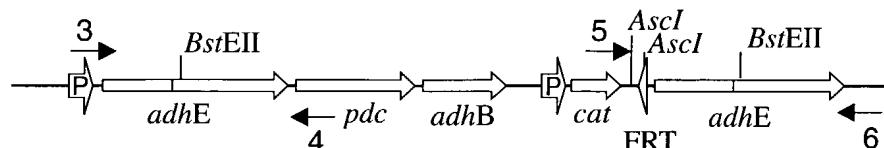


Fig. 3C

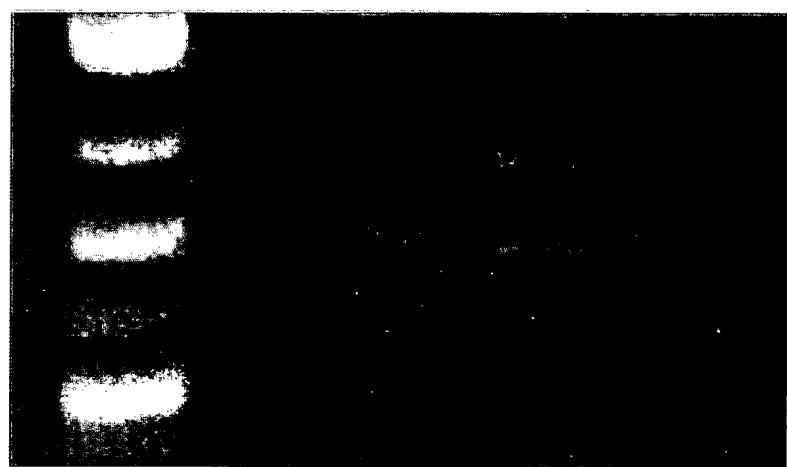


4/4



1 2 3 4 5

Fig. 4A



1 2 3 4 5

Fig. 4B

- 1 -

SEQUENCE LISTING

<110> University of Florida

<120> METHODS AND COMPOSITIONS FOR THE CHROMOSOMAL
INTEGRATION OF HETEROLOGOUS DNA

<130> BCI-021PC

<140>

<141>

<150> 09/390,479

<151> 1999-09-07

<160> 7

<170> PatentIn Ver. 2.0

<210> 1

<211> 2427

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: cloning
vectors

<220>

<223> Ampicillin-resistance gene from position 2295 to
position 1435

<400> 1

atcgatgaat tgatccgaag ttcctattct ctagaaagta taggaacttc gaattgtcga 60

caagcttgat ctggcttatac gaaattaata cgactcacta tagggagacc ggaattccga
120

gcttggcgcg ccaggtcgac tctagaggat ccccggggaa gatcttccgg aagatttcc
180

cgagctcgaa ttaattccgc gatgaattga tccccgaagt tcctattctc tagaaagtat
240

aggaacttcg aattggtcga caagctagct tgcattgcag cttgtattct atagtgac
300

ctaaatcgta tgtgtatgtat acataagggtt atgtatataat tgttagcccg ttctaacgc
360

aatatgtaca agcctaattg tgttagcatct ggcttactga agcagaccct atcatctctc
420

tctgtaaactg ccgtcagagt cggttgggtt ggacgaacct tctgagtttc tggtaacgcc
480

- 2 -

gttccgcacc ccggaaatgg tcagcgaacc aatcagcagg gtcatcgcta gcccatggct
540

aattctgtca gccgttaagt gttcctgtgt cactgaaaat tgcttgaga ggctctaagg
600

gcttctcagt gcgttacatc cctggcttgt tgtccacaac cgtaaacct taaaagctt
660

aaaagcctta tatattcttt ttttcttat aaaacttaaa accttagagg ctatthaagt
720

tgctgattta tattaatttt attgttcaaa catgagagct tagtacgtga aacatgagag
780

cttagtacgt tagccatgag agcttagtac gttagccatg agggtttagt tcgttaaaca
840

tgagagctta gtacgttaaa catgagagct tagtacgtga aacatgagag cttagtacgt
900

actatcaaca ggttgaactg cgatatctgc ggccgaatc gggcaaatcg ctgaatattc
960

cttttgtctc cgaccatcag gcacctgagt cgctgtcttt ttctgtacat tcagttcgct
1020

gcgctcacgg ctctggcagt gaatgggggt aaatggact acaggcgcct tttatggatt
1080

catgcaagga aactcaaaaat aatacaagaa aagccgtca cgggcttctc agggcgttt
1140

atggcgggtc tgctatgtgg tgctatctga cttttgctg ttctgcgtt cctgccctct
1200

gattttccag tctgaccact tcggattatc ccgtgacagg tcattcagac tggctaattgc
1260

acccagtaag gcagcggtat catcaacggg gtctgacgct cagtggAACG aaaactcACG
1320

ttaagggatt ttggcatga gattatcaaa aaggatctc acctagatcc ttttaaattta
1380

aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggtctg acagttacca
1440

atgcttaatc agtgaggcac ctatctcagc gatctgtcta ttctgttcat ccatagttgc
1500

ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc
1560

tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc

- 3 -

1620

agccggaagg gccgagcgca gaagtggtcc tgcaactta tccgcctcca tccagtctat
1680

taattgttgc cgggaagcta gagtaagtag ttccgcagtt aatagttgc gcaacgttgt
1740

tgccattgct acaggcatcg tggtgtcacg ctcgtcgttt ggtatggcatt cattcagctc
1800

cggttcccaa cgatcaaggc gagttacatg atccccatg ttgtgcaaaa aagcggttag
1860

ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt
1920

tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac
1980

tggtgagtagc tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgtcttg
2040

ccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat
2100

tggaaaacgt ttttcggggc gaaaactctc aaggatctt ccgctgttga gatccagttc
2160

gatgtAACCC actcgtgcac ccaactgatc ttcagcatct ttactttca ccagcgttc
2220

tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa
2280

atgttgaata ctcatactct tccttttca atattattga agcatttatac agggttattg
2340

tctcatgagc ggatacatat ttgaatgtat tttagaaaaat aaacaaatag gggttccgcg
2400

cacatTTCCC cgaaaagtgc cacttgc
2427

<210> 2

<211> 1929

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: cloning
vectors

<220>

<223> Kanamycin-resistance gene from position 1778 to

- 4 -

position 984

<400> 2
atcgatgaat tgatccgaag ttccatttct ctagaaagta taggaacttc gaattgtcga 60
caagcttgat ctggcttatac gaaattaata cgactcacta tagggagacc ggaattccga
120
gcttggcgcg ccaggtcgac tctagaggat cccccgggaa gatcttccgg aagatcttcc
180
cgagctcgaa ttaattccgc gatgaattga tcccggaaatg tcctattctc tagaaagtat
240
aggaacttcg aattggtcga caagctagct tgcattgcaag cttgtattct atagtgtcac
300
ctaaatcgta tgtgtatgtat acataaggaaatg atgtattaaat ttagcccg tttctaacgac
360
aatatgtaca agcctaattt ttagcatct ggcttactga agcagaccct atcatctctc
420
tcgttaactg ccgtcagagt cgggttgggtt ggacgaacct tctgagtttc tggtaacgcc
480
gttccgcacc cggaaaatgg tcagcgaacc aatcagcagg gtcattcgcta gcccatggct
540
aattctgtca gcccgttaagt gttccctgtgt cactgaaaat tgctttgaga ggctctaagg
600
gcttctcagt gcgttacatc cctggcttgt tgcacacaac cgttaaacct taaaagcttt
660
aaaaggctta tatattcttt tttttcttat aaaacttaaa accttagagg ctatataatg
720
tgctgattta tattatattt attgttcaaa catgagagct tagtacgtga aacatgagag
780
cttagtacgt tagccatgag agcttagtac gtttagccatg agggtttagt tcgttaaaca
840
tgagagctta gtacgttaaa catgagagct tagtacgtga aacatgagag cttagtacgt
900
actatcaaca ggttgaactg cggatcttgc ggccgcaaaa attaaaaatg aagtttgac
960
ggtatcgaac cccagagtcc cgctcagaag aactcgtcaa gaaggcgata gaaggcgatg
1020
cgctgcgaat cgggagcggc gataccgtaa agcacgagga agcggtcagc ccattcgccg
1080

- 5 -

ccaagctttt cagcaatatac acgggttagcc aacgctatgt cctgatacg gtccgccaca
1140

cccagccggc cacagtcgtt gaatccagaa aagcgccat tttccaccat gatattcggc
1200

aaggcaggcat cgccatgggt cacgacgaga tcctcgccgt cgggcattccg cgccttgagc
1260

ctggcgaaca gttcggctgg cgcgagcccc tggatgtttttt cgtccagatc atcctgatcg
1320

acaagaccgg cttccatccg agtacgtgct cgctcgatgc gatgtttcgc ttgggtggc
1380

aatgggcagg tagccggatc aagcgttatgc agccgcccga ttgcattcagc catgatggat
1440

acttttcggc caggagcaag gtgagatgac aggagatctt gccccggcac ttccggccat
1500

agcagccagt cccttccgc ttcagtgaca acgtcgagca cagctgcgca aggaacgccc
1560

gtcgtggcca gccacgatag ccgcgcgtgcc tcgtcttggg gttcattcag ggcaccggac
1620

aggtcggctt tgacaaaaaag aaccgggcgc ccctgcgctg acagccggaa cacggcggca
1680

tcatagccatcg cgattgtctg ttgtgcccag tcataccatcg atagcctctc caccctcg
1740

gcccggagaac ctgcgtgcaa tccatcttgt tcaatcatgc gaaacgatcc tcattctgtc
1800

tcttgatcca ctagattattttaa gaagcattta tcagggttat tgtctcatga gcggatacat
1860

atttgaatgt attttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt
1920

gccacactgc
1929

<210> 3

<211> 2003

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: cloning
vectors

<220>

- 6 -

<223> chloramphenicol-resistance gene from position 1704
to position 1045

<400> 3

atcgatgaat tgatccgaag ttcctattct ctagaaagta taggaacttc gaattgtcga 60

caagcttgat ctggcttatac gaaattaata cgactcacta tagggagacc ggaattccga
120

gcttggcgcg ccaggtcgac tctagaggat ccccggggaa gatcttccgg aagatcttcc
180

cgagctcgaa ttaattccgc gatgaattga tcccgaaagt tcctattctc tagaaagtat
240

aggaacttcg aattggtcga caagcttagct tgcattgcaag cttgtattct atagtgtcac
300

ctaaatcgta tgtgtatgtat acataaggaa atgtattaaat tgtagcccg ttctaacgac
360

aatatgtaca agcctaattt tgcattcatct ggcttactga agcagaccct atcatctctc
420

tcgttaaactg ccgtcagagt cggtttgggtt ggacgaacct tctgagtttc tggtaacgcc
480

gttccgcacc ccggaaatgg tcagcgaacc aatcagcagg gtcattcgcta gcccatggct
540

aattctgtca gcccgttaagt gttcctgtgt cactgaaaat tgctttgaga ggctctaagg
600

gcttcctcagt gcgttacatc cctggcttgt tgcacacaac cgttaaacct taaaagcttt
660

aaaaggctta tatattcttt tttttcttat aaaacttaaa accttagagg ctatthaagt
720

tgcgttattt tattatattt attgttcaaa catgagagct tagtacgtga aacatgagag
780

cttagtacgt tagccatgag agcttagtac gtttagccatg agggtttagt tcgttaaaca
840

tgagagctta gtacgttaaa catgagagct tagtacgtga aacatgagag cttagtacgt
900

actatcaaca ggttgaactg cggatcttgc ggccgcaaaa attaaaaatg aagttttaaa
960

tcaatctaaa gtatatatga gtaaacttgg tctgacagtt accaatgctt aatcagttag
1020

gcaccaataa ctgccttaaa aaaattacgc cccgcctgc cactcatcgc agtactgttg
1080

- 7 -

taattcatta agcattctgc cgacatggaa gccatcacag acggcatgat gaacctgaat
1140

cgccagcggc atcagcacct tgcgccttg cgtataatat ttgcccattgg tgaaaacggg
1200

ggcgaagaag ttgtccatat tggccacgtt taaatcaaaa ctggtgaaac tcacccaggg
1260

atggctgag acgaaaaaca tattctaat aaaccctta gggaaatagg ccaggtttc
1320

accgtaaacac gccacatott gcgaatataat gtgtagaaac tgccggaaat cgtcggtgt
1380

ttcactccag agcgatgaaa acgtttcagt ttgctcatgg aaaacggtgt aacaagggtg
1440

aacactatcc catatcacca gtcaccgtc tttcattgcc atacggaatt tcggatgagc
1500

attcatcagg cggcaagaa tgtgaataaa ggccggataa aacttgtgt tatTTTctt
1560

tacggcttt aaaaaggccg taatatccag ctgaacggtc tggttatagg tacattgagc
1620

aactgactga aatgcctcaa aatgttcttt acgatgccat tggatataat caacgggt
1680

atatccagt attttttctt ccattttagc ttcccttagct cctgaaaatc tcgataactc
1740

aaaaaatacg cccggtagtg atcttatttc attatggtga aagttggaac ctcttacgtg
1800

ccgatcaacg tctcattttc gccaagggtt ggcccagggtt ttccccgtat caacaggac
1860

accaggattt atttattctg cgaagtgatc ttccgtcaca ggtattttt cggcgaaag
1920

tgcgtcggt gatgctgcca acttactgat ttagtgtatg atgggtttt tgaggtgctc
1980

cagtggcttc tgtttctatc agc
2003

<210> 4
<211> 3246
<212> DNA
<213> Artificial Sequence

<220>

- 8 -

<223> Description of Artificial Sequence: cloning
vectors

<220>

<223> Ampicillin-resistance gene from position 3114 to
position 2254

<400> 4

atcgatgaat tgatccgaag ttcctattct ctagaaagta taggaacttc gaattgtcga 60

caagcttgat ctggcttatac gaaattaata cgactcacta tagggagacc ggaattccga
120

gcttggcgcg ccaggtcgac tctagaggat cccccgggaa gatcttccgg aagatcttcc
180

cgagctcgaa ttaattccgc gatgaattga tcccggaaat tcctattctc tagaaagtat
240

aggaacttcg aattggtcga caagctagct tgcattgcaag ctggcactg gctgatcagc
300

tagcccatgg gtatggacag ttttcccttt gatatgtAAC ggtgaacagt tgttctactt
360

ttgtttgtta gtcttgatgc ttcaactgata gatacaagag ccataagaac ctcagatcct
420

tccgtattta gccagtatgt tctctagtgt ggttcggtgt tttgcgtga gccatgagaa
480

cgaaccattt agatcatgct tactttgcat gtcactcaaa aattttgcct caaaaactgg
540

gagctgaatt tttgcagtta aagcatcgat tagtgtttt cttagtccgt tacgttaggt
600

ggaatctgat gtaatggtg ttggtatttt gtcaccattt attttatct ggttgttctc
660

aagttcggtt acgagatcca tttgtctatc tagttcaact tgaaaaatca acgtatcagt
720

cgggcggcct cgcttatcaa ccaccaattt catattgctg taagtgttta aatctttact
780

tattggtttc aaaaccatt ggttaagcct tttaaactca tggttagttat tttcaagcat
840

taacatgaac ttaaattcat caaggctaat ctctatattt gccttgtgag tttcttttg
900

tgtagttct ttaataacc actcataaat cctcatagag tattttttt caaaagactt
960

aacatgttcc agattatatt ttatgaattt ttttaactgg aaaagataag gcaatatctc

- 9 -

1020

ttcactaaaa actaattcta attttcgct tgagaacttg gcatagttg tccactggaa
1080

aatctcaaag ccttaacca aaggattcct gattccaca gttctcgta tcagctct
1140

ggtgcttta gctaatacac cataagcatt ttccctactg atgttcatca tctgaacgta
1200

tgggtataa gtgaacgata ccgtccgttc ttccttgta gggtttcaa tcgtggggtt
1260

gagtagtgcc acacagcata aaattagctt ggtttcatgc tccgttaagt catagcgact
1320

aatcgctagt tcatttgctt tgaaaacaac taattcagac atacatctca attggcttag
1380

gtgattttaa tcactatacc aattgagatg ggctagtcaa tgataattac tagtccttt
1440

cctttaggtt gtgggtatct gtaaattctg cttagacctt gctggaaaac ttgtaaattc
1500

tgcttagaccc tctgtaaatt ccgctagacc tttgtgtgtt tttttgttt atattcaagt
1560

ggttataatt tatagaataa agaaagaata aaaaaagata aaaagaatag atcccagccc
1620

tgtgtataac tcactacttt agtcagttcc gcagtattac aaaaggatgt cgcaaacgct
1680

gtttgctcct ctacaaaaca gacctaaaaa ccctaaaggc ttaagtagca ccctcgcaag
1740

ctcgggtgcg gccgcaatcg ggcaaatcgc tgaatattcc ttttgtctcc gaccatcagg
1800

cacctgagtc gctgtcttt tcgtgacatt cagttcgctg cgctcacggc tctggcagtg
1860

aatgggggta aatggcacta caggcgccctt ttatggattc atgcaaggaa actaccata
1920

atacaagaaa agcccgtcac gggcttctca gggcgttta tggcggtct gctatgtggt
1980

gctatctgac ttttgctgt tcagcagttc ctgccctctg atttccagt ctgaccactt
2040

cggattatcc cgtgacaggt cattcagact ggctaattgca cccagtaagg cagcggtatc
2100

- 10 -

atcaacgggg tctgacgctc agtggAACGA aaACTCACGT taAGGGATT tggtcatgag
2160

attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagt ttAAATCAAT
2220

ctaaAGTATA tatgagtaaa ctgggtctga cagttaccaa tgcttaatca gtgaggcacc
2280

tatctcagcg atctgtctat ttCGTTcAtC cataGTTGCC tgactccccg tcgtgtAGAT
2340

aactacgata cgggagggct taccatctgg ccccAGTGCT gcaatgatac cgcgagaccc
2400

acgctcacCG gCTCCAGATT tatcAGCAAT aaACCAGCCA GCCGGAAGGG CCGAGCGCAG
2460

aAGTGGTCCT gcaactttat ccgcctccat ccAGTCTATT aATTGTTGCC GGGAAAGCTAG
2520

agtaAGTAGT tcGCCAGTTA atAGTTGCG caACGTTGTT GCCATTGCTA CAGGCATCGT
2580

ggTGTcAcGC tcgtcgTTG gTATGGCTTC ATTcAGCTCC ggTTCCCAAC gATCAAGGCG
2640

agttacatga tccccatgt tgtcaaaaaa agcggtagc tccttcggc tcCcGATCGT
2700

tgtcagaAGT aAGTGGCCG cAGTGTtACt ACTCATGGTT ATGGCAGCAC TGCATAATTc
2760

tcttactgtc atGCCATCCG taAGATGCTT ttctgtgact ggtgAGTACT caACCAAGTC
2820

attctgagaa tagtgtatgc ggCGACCGAG ttGCTCTTGC CCGGCGTCAA tacGGATAA
2880

taccGCGCCA catAGCAGAA ctTTAAAAGT gCTCATCATT ggAAAACGTT cttcGGGGCG
2940

aaaACTCTCA aggatcttac CGCTGTTGAG ATCCAGTTG ATGTAACCCA CTcGTGcAcc
3000

caactgatct tcAGCATCTT ttACTTtCAC cAGCgtttct gggtgAGCAA AACAGGAAAG
3060

gcaAAATGCC gcaAAAAAGG gaATAAGGGC gACACGGAAA tGTTGAATAC tcataCTT
3120

ccttttcaa tattattgaa gcatttatca gggTTATTGT cTCATGAGCG gatacatATT
3180

tGAATGTATT tagAAAAATA aACAAATAGG ggTCCGCGC ACATTCCCC gAAAAGTGC
3240

- 11 -

acttg
3246

<210> 5
<211> 3443
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: cloning
vectors

<220>
<223> kanamycin-resistance gene from position 3292 to
position 2498

<400> 5
atcgatgaat tgatccgaag ttcctattct ctagaaagta taggaacttc gaattgtcga 60

caagcttgcgat ctggcttatac gaaattaata cgactcacta tagggagacc ggaattccga
120

gcttggcgcg ccaggtcgac tctagaggat cccccgggaa gatcttccgg aagatcttcc
180

cgagctcgaa ttaattccgc gatgaattga tcccggaaatg tcctattctc tagaaagtat
240

aggaacttcg aattggtcga caagcttagct tgcatgcaag ctggcactg gctgatcagc
300

tagccatgg gtatggacag tttcccttt gatatgtaac ggtgaacagt tgttctactt
360

ttgtttgtta gtcttgatgc ttcaactgata gatacaagag ccataagaac ctcagatcct
420

tccgtattta gccagtatgt tctctagtgt ggttcggtgt tttgcgtga gccatgagaa
480

cgaaccattg agatcatgct tactttgcat gtcactcaa aattttgcct caaaaactgg
540

gagctgaatt tttgcagtta aagcatcggt tagtgtttt cttagtccgt tacgttaggt
600

ggaatctgat gtaatgggtt ttggtatattt gtcaccattc attttatct ggttgttctc
660

aagttcggtt acgagatcca tttgtctatc tagttcaact tgaaaaatca acgtatcagt
720

ccccccgcct cgcttatcaa ccaccaattt catattgctg taagtgttta aatcttact
780

- 12 -

tattgtttc aaaacccatt ggttaagcct tttaaactca tggttagttat tttcaagcat
840

taacatgaac ttaaattcat caaggctaat ctctatattt gccttgtgag tttcttttg
900

tgttagttct tttaataacc actcataaaat cctcatagag tatttggggaaaagactt
960

aacatgttcc agattatattt ttatgaattt ttttaactgg aaaagataag gcaatatctc
1020

ttcactaaaa actaattcta attttcgct tgagaacttg gcatacggttgcactggaa
1080

aatctcaaag ctttaacca aaggattcct gatttccaca gttctcgta tcagctctt
1140

ggttgtttaa gtgaacgata ccgtccgttc ttccctgtt gggtttcaa tcgtggggtt
1200

gagtagtgcc acacagcata aaatttagctt ggtttcatgc tccgttaagt catagcgact
1320

aatcgctagt tcatttgctt tgaaaacaac taattcagac atacatctca attggcttag
1380

gtgattttaa tcactatacc aattgagatg ggctagtcaa tgataattac tagtccttt
1440

cctttgagtt gtgggtatct gtaaattctg ctagaccttt gctggaaaac ttgttaattc
1500

tgcttagaccc tctgtaaatt ccgttagacc tttgtgtgtt tttttgttt atattcaagt
1560

ggttataatt tatagaataa agaaagaata aaaaaagata aaaagaatag atcccagccc
1620

tgtgtataac tcactacttt agtcagttcc gcagtattac aaaaggatgt cgcaaacgct
1680

gtttgctcct ctacaaaaca gacctaaaaa ccctaaaggc ttaagtagca ccctcgcaag
1740

ctcggttgcg gccgcaatcg ggcaaattcgc tgaatattcc ttttgcgttcc gaccatcagg
1800

cacctgagtc gctgtttt tcgtgacatt cagttcgctg cgctcacggc tctggcagtg
1860

aatgggggta aatggcacta caggcgccctt ttatggattc atgcaaggaa actacccata

- 13 -

1920

ataacaagaaa agcccgac gggcttctca gggcggttta tggcgggtct gctatgtggt
1980

gctatctgac ttttgctgt tcagcagttc ctgcctctg atttccagt ctgaccactt
2040

cggattatcc cgtgacaggt cattcagact ggctaattgca cccagtaagg cagcggtatc
2100

atcaacgggg tctgacgctc agtggAACGA aaactcacgt taaggattt tggtcatgag
2160

attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttAAATCAAT
2220

ctaaagtata tatgagtaaa cttggctctga cagttaccaa tgcttaatca gtgaggcacc
2280

tatctcagcg atctgtctat ttCGTTcatc catagttgcc tgactccccg tcgtgttagat
2340

aactacgata cgggagggct taccatctgg ccccagtgtc gcaatgatac cgcgagaccc
2400

acgctcaccg gctccagatt tatcagcaat aaaccagcca gccgggcccgc aaaaattaaa
2460

aatgaagttt tgacggtatc gaaccccaga gtcccgctca gaagaactcg tcaagaaggc
2520

gatagaaggc gatgcgctgc gaatcgggag cggcgatacc gtaaagcacg aggaagcggt
2580

cagcccatc gccgccaagc tcttcagcaa tatcacgggt agccaacgct atgtcctgat
2640

agcggtccgc cacacccagc cggccacagt cgatgaatcc agaaaaagcgg ccattttcca
2700

ccatgatatt cggcaaggcag gcatcgccat gggtcacgac gagatcctcg ccgtcgggca
2760

tccgcgcctt gagcctggcg aacagttcgg ctggcgcgag cccctgatgc tcttcgtcca
2820

gatcatcctg atcgacaaga ccggcttcca tccgagtacg tgctcgctcg atgcgatgtt
2880

tcgcttggtg gtcgaatggg caggtagccg gatcaagcgt atgcagccgc cgcattgcat
2940

cagccatgat ggatactttc tcggcaggag caaggtgaga tgacaggaga tcctgccccg
3000

- 14 -

gcacttcgcc caatagcagc cagtcccttc ccgcttcagt gacaacgtcg agcacagctg
3060

cgcaaggaac gcccgtcgtg gccagccacg atagccgcgc tgccctcgctc tggagttcat
3120

tcagggcacc ggacaggtcg gtcttgacaa aaagaaccgg gcgcacctgc gctgacagcc
3180

ggaacacggc ggcattcagag cagccgattt tctgttgtgc ccagtcatacg ccgaatagcc
3240

tctccaccca agcggccgga gaacctgcgt gcaatccatc ttgttcaatc atgcgaaacg
3300

atcctcatcc tgtctcttga tccacttagat tattgaagca tttatcaggg ttattgtctc
3360

atgagcggat acatatttga atgtatTTAG aaaaataaaac aaataggGGT tccgcgcaca
3420

tttccccgaa aagtgccacc tgc
3443

<210> 6

<211> 3508

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: cloning
vectors

<220>

<223> chloramphenicol-resistance gene from position 3209
to position 2550

<400> 6

atcgatgaat tgatccgaag ttcctattct ctagaaagta taggaacttc gaattgtcga 60

caagcttcat ctggcttatac gaaattaata cgactcacta tagggagacc ggaattccga
120

gcttggcgcg ccaggtcgac tctagaggat ccccgaaaa gatcttcgg aagatttcc
180

cgagctcgaa ttaattccgc gatgaattga tcccgaaatg tcctattctc tagaaagtat
240

aggaacttcg aattggtcga caagctagct tgcatgcaag cttggcactg gctgatcagc
300

tagccatgg gtatggacag tttcccttt gatatgtaac ggtgaacagt tgttctactt
360

- 15 -

ttgttgttta gtcttgatgc ttcaactgata gataacaagag ccataagaac ctcagatcct
420

tccgtattta gccagtatgt tctctagtgt ggttcggtgt ttttgcgtga gccatgagaa
480

cgaaccattg agatcatgct tactttgcat gtcactcaaa aattttgcct caaaaactgg
540

gagctgaatt tttgcagtttta aagcatcggt tagtgttttt cttagtccgt tacgttaggt
600

ggaatctgat gtaatggttg ttggattttt gtcaccattc attttatct gggttgc
660

aagttcgggtt acgagatcca tttgtctatc tagttcaact tgaaaaatca acgtatcgt
720

cggggcggcct cgcttatcaa ccaccaattt catattgctg taagtgttta aatctttact
780

tattggtttc aaaaccatt ggttaaggcct tttaaactca tggtagttat tttcaagcat
840

taacatgaac ttaaattcat caaggctaat ctctatattt gccttgcgtt gtttttttgc
900

tgttagttct ttaataacc actcataaaat cctcatagag tatttggttt caaaagactt
960

aacatgttcc agattatattt ttatgaattt ttttaactgg aaaagataag gcaatatctc
1020

ttcactaaaa actaattcta attttcgct tgagaacttg gcatagtttgc tccactggaa
1080

aatctcaaag ccttaacca aaggattcct gattccaca gttctcgatca tcagctctct
1140

ggttgcttta gctaatacac cataagcatt ttccctactg atgttcatca tctgaacgta
1200

ttgggtataa gtgaacgata ccgtccgttc ttcccttgta gggttttcaa tcgtggggtt
1260

gagtagtgcc acacagcata aaatttagctt ggtttcatgc tccgttaagt catagcgact
1320

aatcgctagt tcatttgctt tgaaaacaac taattcagac atacatctca attggcttag
1380

gtgattttaa tcactatacc aattgagatg ggcttagtcaa tgataattac tagtccttt
1440

cctttgagtt gtgggtatct gtaaattctg ctagaccttt gctggaaaac ttgtaaattc
1500

- 16 -

tgctagaccc tctgtaaatt ccgctagacc tttgtgtgtt tttttgttt atattcaagt
1560

ggttataatt tatagaataa agaaagaata aaaaaagata aaaagaatag atcccagccc
1620

tgtgtataac tcactacttt agtcagttcc gcagtattac aaaaggatgt cgcaaacgct
1680

gtttgctcct ctacaaaaca gaccttaaaa ccctaaaggc ttaagtagca ccctcgcaag
1740

ctcggttgcg gccgcaatcg ggcaaatcgc tgaatattcc tttgtctcc gaccatcagg
1800

cacctgagtc gctgtcttt tcgtgacatt cagttcgctg cgctcacggc tctggcagtg
1860

aatggggta aatggcacta caggcgctt ttatggattc atgcaaggaa actacccata
1920

atacaagaaa agcccgtcac gggcttctca gggcgttta tggcggtct gctatgtgg
1980

gctatctgac ttttgctgt tcagcagttc ctgccctctg atttccagt ctgaccactt
2040

cgattatcc cgtgacaggt cattcagact ggctaattgca cccagtaagg cagcggtata
2100

atcaacgggg tctgacgctc agtggAACGA aaactcacgt taagggattt tggcatgag
2160

attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttAAATCAAT
2220

ctaaagtata tatgagtaaa cttggctga cagttacca tgcttaatca gtgaggcacc
2280

tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgttagat
2340

aactacgata cgggagggt taccatctgg ccccagtgt gcaatgatac cgcgagaccc
2400

acgctcaccc gctccagatt tatcagcaat aaaccagcca gccgaattaa aaatgaagtt
2460

ttaaatcaat ctaaagtata tatgagtaaa cttggctga cagttacca tgcttaatca
2520

tgaggcacc aataactgcc ttaaaaaaaaat tacgccccgc cctgccactc atcgcagtg
2580

tgttgtaatt cattaagcat tctgccgaca tggaagccat cacagacggc atgatgaacc

- 17 -

2640

tgaatcgcca gcggcatcag caccttgcg ccttgcgtat aatatttgcc catggtaaaa
2700

acggggggcga agaagttgtc catattggcc acgtttaaat caaaaactggt gaaactcacc
2760

cagggattgg ctgagacgaa aaacatattc tcaataaacc cttagggaa ataggccagg
2820

ttttcacccgt aacacgcccac atcttgcgaa tataatgtgt aaaaactgccc gaaatcgctcg
2880

tggtatttcac tccagagcga tgaaaacgtt tcagtttgct catggaaaac ggtgtaacaa
2940

gggtgaacac tatccccatat caccagctca ccgtctttca ttgccatacg gaatttcgga
3000

tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaaactt gtgcttattt
3060

ttctttacgg tctttaaaaaa ggccgtaata tccagctgaa cggctctgggt ataggtacat
3120

tgagcaactg actgaaaatgc ctcaaaatgt tctttacgat gccattggga tataatcaacg
3180

gtggtatatac cagtgatttt tttctccatt ttagcttcct tagctcctga aaatctcgat
3240

aactcaaaaaa atacgccccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt
3300

acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca
3360

gggacaccag gatttattta ttctgcgaag tgatcttcgg tcacaggtat ttattcggcg
3420

caaagtgcgt cgggtgatgc tgccaaactta ctgatttagt gtatgatggt gttttgagg
3480

tgctccagtg gcttctgttt ctatcagc
3508

<210> 7

<211> 9460

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: cloning
vectors

- 18 -

<220>
<223> adhE gene from E.coli from position 1890 to position 4565

<220>
<223> pdc gene from Z.mobilis from position 4701 to position 6380

<220>
<223> adhB gene from Z.mobilis from position 6508 to position 7659

<220>
<223> chloramphenicol-resistance gene from position 8515 to position 9174

<220>
<223> kanamycin-resistance gene from position 152 to position 946

<400> 7
gcaggtggca cttttcgggg aaatgtgcgc ggaaccctta ttgtttatt tttctaaata 60
cattcaaata tgtatccgct catgagacaa taaccctgat aaatgcttca ataatctagt
120
ggatcaagag acaggatgag gatcgttcg catgattgaa caagatggat tgcacgcagg
180
ttctccggcc gcttgggtgg agaggctatt cggctatgac tgggcacaac agacaatcgg
240
ctgctctgat gccgcccgtgt tccggctgtc agcgcagggg cgccccgttc ttttgtcaa
300
gaccgacctg tccggtgccc tgaatgaact ccaagacgag gcagcgcggc tatcgtggct
360
ggccacgacg ggcgttcctt gcgcagctgt gctcgacggt gtcaactgaag cggaaaggga
420
ctggctgcta ttgggcgaag tgccggggca ggatctcctg tcatctcacc ttgctcctgc
480
cgagaaaagta tccatcatgg ctgatgcaat gcggcggctg catacgcttg atccggctac
540
ctgcccattc gaccaccaag cgaaacatcg catcgagcga gcacgtactc ggtatggaagc
600

- 19 -

cggcttgtc gatcaggatg atctggacga agagcatcg gggctcgcc cagccgaact
660
gttcgccagg ctcaaggcgc ggatgcccga cggcgaggat ctcgtcgtga cccatggcga
720
tgccctgcttgc cgaatatca tggtgaaaaa tggccgcttt tctggattca tcgactgtgg
780
ccggctgggt gtggcggacc gctatcagga catagcgttg gctaccctg atattgctga
840
agagcttggc ggcgaatggg ctgaccgctt ctcgtgctt tacggtatcg ccgctcccga
900
ttcgcagcgc atcgcccttct atcgccctct tgacgagttc ttctgagcgg gactctgggg
960
ttcgataccg tcaaaacttc atttttaatt tttgcggccg caagatccgc agttcaacct
1020
gttgatagta cgtactaagc tctcatgttt cacgtactaa gctctcatgt ttaacgtact
1080
aagctctcat gtttaacgaa ctaaacccctc atggctaaca tactaagctc tcatggctaa
1140
cgtactaagc tctcatgttt cacgtactaa gctctcatgt ttgaacaata aaattaatat
1200
aaatcagcaa cttaaatagc ctctaagggtt ttaagttta taagaaaaaa aagaatataat
1260
aaggctttta aagctttaa gtttaacgg ttgtggacaa caagccaggg atgtaacgca
1320
ctgagaagcc ctttagagcct ctcaaagcaa ttttcagtga cacaggaaca cttaacggct
1380
gacagaatta gccatgggtt agcgatgacc ctgctgattt gttcgctgac catttccggg
1440

- 20 -

gtgcggaacg gcgttaccag aaactcagaa ggtcgtcca accaaaccga ctctgacggc
1500
agtttacgag agagatgata gggctcgctt cagtaagcca gatgctacac aattaggctt
1560
gtacatatgg tcgtagaaac gcggctacaa ttaatacata accttatgta tcatacacat
1620
acgattnagg tgacactata gaatacaagc ttgcattgcaa gctagcttgt cgaccaattc
1680
gagttccat actttctaga gaataggaac ttccgggatc aattcatcgc ggaattaatt
1740
cgagctcggg aagatcttcc ggaagatctt ccccgggat cctctagagt cgacctggcg
1800
cgcccccttga ctagagggtc gacgcattgcc tgtacatccg gagacgcgtc acggccgaag
1860
ctagcgaatt cgcccttaat tgctttcca tggctgttac taatgtcgct gaacttaacg
1920
caactcgtaga gcgtgtaaaa aaagcccagc gtgaatatgc cagtttcaact caagagcaag
1980
tagacaaaat cttccgcgcc gcccgtctgg ctgctgcaga tgctcgaatc ccactcgcg
2040
aaatggccgt tgccgaatcc ggcattggta tcgtcgaaga taaagtgtatc aaaaaccact
2100
ttgcttctga atatatctac aacgcctata aagatgaaaa aacctgtggt gttctgtctg
2160
aagacgacac ttttggtacc atcactatcg ctgaaccaat cggtattatt tgcggtatcg
2220
ttccgaccac taacccgact tcaactgcta tcttcaaatac gctgatcagt ctgaagaccc
2280

- 21 -

gtaacgccat tatcttctcc ccgcacccgc gtgcaaaaga tgccaccaac aaagcggctg

2340

atatcggtct gcaggctgct atcgctgccg gtgctccgaa agatctgatc ggctggatcg

2400

atcaacacctc tggtaactg tctaacgcac tggatgcacca cccagacatc aacctgatcc

2460

tgcgcactgg tggccgggc atggtaaag ccgcatacag ctccggtaaa ccagctatcg

2520

gtgttaggcgc gggcaacact ccagttgtta tcgatgaaac tgctgatatc aaacgtgcag

2580

ttgcacatctgt actgatgtcc aaaacacctcg acaacggcgt aatctgtgct tctgaacagt

2640

ctgttgttgt tggtaactct gtttatgacg ctgtacgtga acgttttgca acccacggcg

2700

gctatctgtt gcagggtaaa gagctgaaag ctgttcagga tggtaatcctg aaaaacggtg

2760

cgctgaacgc ggctatcggtt ggtcagccag cctataaaat tgctgaactg gcaggcttct

2820

ctgtaccaga aaacacccaag attctgatcg gtgaagtgac cgttgttgat gaaagcgaac

2880

cgttcgcaca tggaaaaactg tccccgactc tggcaatgta ccgcgctaaa gatttcgaag

2940

acgcggtaga aaaagcagag aaactggttg ctatggcgg tatcggtcat acctcttgc

3000

tgtacactga ccaggataac caaccggctc gcgtttctta ctgcggtcag aaaatgaaaa

3060

cggcgcgtat cctgattaac accccagcgt ctcagggtgg tatcggtgac ctgtataact

3120

- 22 -

tcaaactcgacacctccatg actctgggtt gtggttcttg gggtggtAAC tccatctctg
3180
aaaacgttgg tccgaaacac ctgatcaaca agaaaaccgt tgctaagcga gctgaaaaca
3240
tgttgtggca caaacttccg aaatcttatct acttccgccc tggctccctg ccaatcgcc
3300
tggatgaagt gattactgat ggccacaaac gtgcgcgtcat cgtgactgac cgcttcctgt
3360
tcaacaatgg ttatgctgat cagatcactt ccgtactgaa agcagcaggc gttgaaactg
3420
aagtcttctt cgaagtagaa gcggacccga ccctgagcat cgttcgtaaa ggtgcagaac
3480
tggcaaactc cttcaaacca gacgtgatta tcgcgcgtgg tggtggttcc ccgatggacg
3540
ccgcgaagat catgtgggtt atgtacgaac atccggaaac tcacttcgaa gagctggcgc
3600
tgcgctttat ggatatccgt aaacgttatct acaagttccc gaaaatgggc gtgaaagcga
3660
aatgatcgc tgtcaccacc acttctggta caggttctga agtcactccg tttgcgggtt
3720
taactgacga cgctactggc cagaaatatac cgctggcaga ctatgcgcgt actccggata
3780
tggcgattgt cgacgccaac ctggttatgg acatgccaa gtcctgtgt gcttcgggt
3840
gtctggacgc agtaactcac gccatggaaag cttatgttcc tgtactggca tctgagttct
3900
ctgatggtca ggctctgcag gcactgaaac tgctgaaaga atatctgcca gcgtcctacc
3960

- 23 -

acgaagggtc taaaaatccg gtagcgcgtg aacgtgttca cagtgcagcg actatgcgg

4020

gtatcgcgtt tgcaacgcc ttctgggtg tatgtcactc aatggcgac aaactgggtt

4080

cccagttcca tattccgcac ggtctggcaa acgcccgtct gattttaac gttattcgct

4140

acaatgcgaa cgacaacccg accaagcaga ctgcatttag ccagtatgac cgtccgcagg

4200

ctcgccgtcg ttatgctgaa attgccgacc acttgggtct gagcgcaccc ggccgaccgta

4260

ctgctgctaa gatcgagaaa ctgctggcat ggctggaaac gctgaaagct gaactgggtt

4320

ttccgaaatc tatccgtgaa gctggcgttc aggaagcaga cttcctggcg aacgtggata

4380

aactgtctga agatgcattc gatgaccagt gcaccggcgc taacccgctg taccggctga

4440

tctccgagct gaaacagatt ctgctggata cctactacgg tcgtgattat gtagaaggtg

4500

aaactgcagc gaagaaagaa gctgctccgg ctaaagctga gaaaaaaagcg aaaaaaatccg

4560

cttaacgaag agcaattaag ggcgaattcg tggatcctcg tcggttcaaa aaatgcctat

4620

agctaaatcc ggaacgacac ttttagaggtt tctgggtcat cctgattcag acatagtgtt

4680

ttgaatatat ggagtaagca atgagttata ctgtcggtac ctattttagcg gcgcttgcc

4740

agattgggtct caagcatcac ttgcgcgtcg cgggcgacta caacctcgtc cttcttgaca

4800

- 24 -

acctgctttt gaacaaaaac atggagcagg tttattgctg taacgaactg aactgcgggtt

4860

tcaagtgcaga aggttatgct cgtccaaag cggacgcagc agccgtcggtt acctacagcg

4920

tcggtgcgct ttccgcattt gatgctatcg gtggcgcccta tgcagaaaac cttccggta

4980

tcctgatctc cggtgctccg aacaacaatg atcacgctgc tggcacgtg ttgcacgtc

5040

ctcttggcaa aaccgactat cactatcagt tggaaatggc caagaacatc acggccgcag

5100

ctgaagcgat ttacacccca gaagaagctc cggctaaaat cgatcacgtg attaaaactg

5160

ctcttcgtga gaagaagccg gtttatctcg aaatcgcttg caacattgct tccatgccct

5220

gccccgtcc tggaccggca agcgcattgt tcaatgacga agccagcgac gaagcttctt

5280

tgaatgcagc gggtgaagaa accctgaaat tcatcgccaa ccgcgacaaa gttgccgtcc

5340

tcgtcggcag caagctgcgc gcagctggtg ctgaagaagc tgctgtcaaa tttgctgatg

5400

ctctcggtgg cgcagttgct accatggctg ctgcaaaaag cttcttccag aagaaaaccg

5460

cattacatcg gtacctcatg ggtgaagtca gctatccggg cgttggaaag acgatgaaag

5520

aagccgatgc ggttatcgct ctggctcctg tcttcaacga ctactccacc actggttgga

5580

cggatattcc tgatcctaag aaactggttc tcgctgaacc gcgttctgtc gtcgtaacg

5640

- 25 -

gcgttcgctt ccccagcgtt catctgaaag actatctgac ccgtttggct cagaaagttt

5700

ccaagaaaac cggtgctttg gacttcttca aatccctcaa tgcaggtgaa ctgaagaaag

5760

ccgctccggc tgatccgagt gctccgttgg tcaacgcaga aatcgccccgt caggtcgaag

5820

ctcttctgac cccgaacacg acggttatttgc ctgaaaccgg tgactcttgg ttcaatgctc

5880

agcgcatgaa gctcccgaac ggtgctcgcg ttgaatatga aatgcagtgg ggtcacatcg

5940

gttggtccgt tcctgccgcc ttcggttatg ccgtcggtgc tccggAACGT cgcaacatcc

6000

tcatggttgg tgatggttcc ttccagctga cggctcagga agtcgctcag atggttcgcc

6060

tgaaactgcc ggttatcatc ttcttgatca ataactatgg ttacaccatc gaagttatga

6120

tccatgatgg tccgtacaac aacatcaaga actgggatta tgccggctcg atgaaagtgt

6180

tcaacggtaa cggtggttat gacagcggcg ctggtaaagg cctgaaggct aaaaccggtg

6240

gcgaactggc agaagctatc aagggttgc tggcaaacac cgacggccca accctgatcg

6300

aatgcttcat cggtcgtgaa gactgcactg aagaatttgtt caaatgggtt aagcgcgttg

6360

ctgcccccca acagccgtaa gcctgttaac aagctcctct agtttttaaa acgggattaa

6420

aacgcaaaaaa caatagaaag cgatttctcg aaaatgggtt ttttcgggtt gttgctttaa

6480

- 26 -

actagtatgt agggtgaggt tatacgatg gcttcttcaa cttttatat tccttcgtc

6540

aacgaaatgg gcgaagggttc gcttgaaaaa gcaatcaagg atcttaacgg cagcggctt

6600

aaaaatgcgc tgatcgtttc tgatgctttc atgaacaaat ccggtgttgt gaagcagggtt

6660

gctgacctgt tgaaaggcaca gggattttat tctgctgttt atgatggcgt tatgccgaac

6720

ccgactgtta ccgcagttct ggaaggcctt aagatcctga aggataacaa ttcagacttc

6780

gtcatctccc tcggtggtgg ttctccccat gactgcgc aagccatcgc tctggtcga

6840

accaatggtg gtgaagtcaa agactacgaa ggtatcgaca aatctaagaa acctgcctg

6900

ccttgatgt caatcaacac gacggctggt acggcttctg aaatgacgac tttctgcac

6960

atcactgatg aagtccgtca cgttaagatg gccattgttg accgtcacgt taccccgatg

7020

gttccgtca acgatcctct gttgatggtt ggtatgc当地 aaggcctgac cgccgccacc

7080

ggtatggatg ctctgaccca cgcatttcaa gcttattctt caacggcagc tactccgatc

7140

accgatgctt gcgcttgaa agcagcttcc atgatcgcta agaatctgaa gaccgcttgc

7200

gacaacggta aggatatgcc ggctcgtgaa gctatggctt atgcccatt cctcgctgg

7260

atggccttca acaacgcttc gcttggttat gtccatgcta tggctcacca gttggggcggt

7320

- 27 -

tactacaacc tgccgcattgg tgtctgcaac gctgttctgc ttccgcattgt tctggcttat

7380

aacgcctctg tcgttgctgg tcgtctgaaa gacgttggtg ttgctatggg tctcgatatac

7440

gccaatctcg gtgataaaga aggcgcagaa gccaccattc aggctgttcg cgatctggct

7500

gctcccattg gtattccagc aaacctgacc gagctgggtg ctaagaaaga agatgtgccg

7560

cttcttgctg accacgctct gaaagatgct tgtgctctga ccaacccgcg tcagggtgat

7620

cagaaagaag ttgaagaact cttcctgagc gctttctaattt ttc当地acag gaaaacggtt

7680

ttccgtcctg tcttgattttt caagcaaaca atgcctccga tttctaattcg gaggcatttg

7740

tttttgttta ttgcaaaaac aaaaaatattt gttacaaattt tttacaggct attaaggcta

7800

ccgtcataaaa taatttgcca tttaaaggct attatcagga tttcgcccc gatttcagcc

7860

atgtatcgct catgatcgcg acatgttctg atatttcct ctaaaaaaga taaaaagtct

7920

tttcgttctg gcagaagagg ttcatcatga acaaaaaattc ggcatttttg atgtaaccca

7980

ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgttctt gggtgagcaa

8040

aaacaggaag gcaaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tggtaataac

8100

tcataactctt ctttttcaa tattattgaa gcatttatca gggattttgt ctcatgagcg

8160

- 28 -

gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc

8220

gaaaagtgcc acctgacgac taagaaacca ttattatcat gacattaacc tataaaaata

8280

ggcgtatcac gaggccctt cgtctcgaa taaatacctg tgacggaaga tcacttcgca

8340

gaataaataa atcctggtgt ccctgttgat accgggaagc cctgggccaat cttttggcga

8400

aatgagacg ttgatcggca cgtaagaggt tccaacttc accataatga aataagatca

8460

ctaccggcg tatttttga gttatcgaga ttttcaggag ctaaggaagc taaaatggag

8520

aaaaaaaaatca ctggatatac caccgttgat atatccaat ggcacatcgtaa agaacatTTT

8580

gaggcatttc agtcagttgc tcaatgtacc tataaccaga ccgttcagct ggatattacg

8640

gccttttaa agaccgtaaa gaaaaataag cacaagttt atccggcctt tattcacatt

8700

cttgcggcc tcatgtatgc tcattccggaa ttccgtatgg caatgaaaga cggtgagctg

8760

gtgatatggg atagtgttca ccctgtttac accgtttcc atgagcaaac tgaaacgttt

8820

tcatcgctct ggagtgaata ccacgacgat ttccggcagt ttctacacat atattcgaa

8880

gatgtggcgt gttacgggtga aaacctggcc tatttcctta aagggtttat tgagaatatg

8940

tttttcgtct cagccaatcc ctgggtgagt ttcaccagtt ttgatttaaa cgtggccaat

9000

- 29 -

atggacaact tcttcgcccc cgtttcacc atgggcaaat attatacgca aggcgacaag

9060

gtgctgatgc cgctggcgat tcaggttcat catgccgtt gtgatggctt ccatgtcggc

9120

agaatgctta atgaattaca acagtactgc gatgagtggc agggcggggc gtaattttt

9180

taaggcagtt attggtgccc ttaaacgcct ggtgctacgc ctgaataagt gataataagc

9240

ggatgaatgg cagaaattcg aaagcaaatt cgaccggat ccagatatcc tgcagagaag

9300

cttggcgcca gccggcttca attgcacggg cgcgccaagc tcggaattcc ggtctcccta

9360

tagtgagtcg tattaatttc gataagccag atcaagcttg tcgacaattc gaagttcccta

9420

tactttctag agaataggaa cttcggatca attcatcgat

9460

INTERNATIONAL SEARCH REPORT

Inte	onal Application No
PCT/US 00/22700	

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/85 C12N15/90 C12N1/21 //C12N15/52,C12N15/52,
(C12N1/21,C12R1:19)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 99 10488 A (WAHL GEOFFREY ;GORMAN STEPHEN O (US); SALK INST FOR BIOLOGICAL STU) 4 March 1999 (1999-03-04)</p> <p>examples 1,3,5 figure 1</p> <p>----</p> <p>-/-</p>	<p>1,4-17, 96,99, 100,102, 106-110, 112-115, 117</p>

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

5 January 2001

Date of mailing of the international search report

23/01/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340 2046

Authorized officer

Galli, I

INTERNATIONAL SEARCH REPORT

Int'l Application No

PCT/US 00/22700

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HASAN N ET AL: "ESCHERICHIA COLI GENOME TARGETING, I. CRE-LOX-MEDIATED IN VITRO GENERATION OF ORI-PLASMIDS AND THEIR IN VIVO CHROMOSOMAL INTEGRATION AND RETRIEVAL" GENE, NL, ELSEVIER BIOMEDICAL PRESS. AMSTERDAM, vol. 150, no. 1, 2 December 1994 (1994-12-02), pages 51-56, XP002916292 ISSN: 0378-1119 cited in the application the whole document	1-17, 21-27, 96-102, 106-115, 117
Y	---	18-20, 28-95, 103-105, 116
Y	WO 92 10561 A (UNIV FLORIDA) 25 June 1992 (1992-06-25)	18-20, 28-95, 103-105, 116
A	examples 3-6 figure 2 --- POSFAI G. ET AL.: "Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion and rescue of the pathogenicity island LEE of the Escherichia coli O157:H7 genome." J. BACTERIOL., vol. 179, no. 13, July 1997 (1997-07), pages 4426-4428, XP002156608 figure 2	1-117
P, X	--- MARTINEZ-MORALES F. ET AL.: "Chromosomal integration of heterologous DNA in Escherichia coli with precise removal of markers and replicons used during construction" J. BACTERIOL., vol. 181, no. 22, November 1999 (1999-11), pages 7143-7148, XP002156609 the whole document	1-117

INTERNATIONAL SEARCH REPORT
Information on patent family members

Int'l Application No

PCT/US 00/22700

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 9910488	A	04-03-1999	AU	9209098 A		16-03-1999
WO 9210561	A	25-06-1992	AT	177784 T		15-04-1999
			AU	1858699 A		09-09-1999
			AU	694674 B		23-07-1998
			AU	6194696 A		31-10-1996
			AU	668677 B		16-05-1996
			AU	9098891 A		08-07-1992
			BR	9107210 A		08-02-1994
			CA	2097803 A		08-06-1992
			DE	69131015 D		22-04-1999
			DE	69131015 T		18-11-1999
			DK	560885 T		11-10-1999
			EP	0560885 A		22-09-1993
			ES	2132116 T		16-08-1999
			FI	932559 A		27-07-1993
			GR	3030570 T		29-10-1999
			JP	6504436 T		26-05-1994
			US	6107093 A		22-08-2000
			US	5482846 A		09-01-1996
			US	5821093 A		13-10-1998
			US	5916787 A		29-06-1999